



# When does OMP achieve exact recovery with continuous dictionaries?

Clément Elvira, Rémi Gribonval, Charles Soussen, Cédric Herzet

## ► To cite this version:

Clément Elvira, Rémi Gribonval, Charles Soussen, Cédric Herzet. When does OMP achieve exact recovery with continuous dictionaries?. *Applied and Computational Harmonic Analysis*, 2021, 51, pp.39. 10.1016/j.acha.2020.12.002 . hal-02099464v4

**HAL Id: hal-02099464**

**<https://inria.hal.science/hal-02099464v4>**

Submitted on 21 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# When does OMP achieve exact recovery with continuous dictionaries?

Clément Elvira<sup>1</sup>, Rémi Gribonval<sup>1</sup>, Charles Soussen<sup>2</sup> and Cédric Herzet<sup>1</sup>

<sup>1</sup> Univ Rennes, Inria, CNRS, IRISA  
F-35000 Rennes, France

<sup>2</sup> L2S, CentraleSupélec-CNRS-Université Paris-Saclay  
91192 Gif-sur-Yvette, France

---

## Abstract

This paper presents new theoretical results on sparse recovery guarantees for a greedy algorithm, Orthogonal Matching Pursuit (OMP), in the context of continuous parametric dictionaries. Here, the continuous setting means that the dictionary is made up of an infinite uncountable number of atoms. In this work, we rely on the Hilbert structure of the observation space to express our recovery results as a property of the kernel defined by the inner product between two atoms. Using a continuous extension of Tropp's Exact Recovery Condition, we identify key assumptions allowing to analyze OMP in the continuous setting. Under these assumptions, OMP unambiguously identifies in exactly  $k$  steps the atom parameters from any observed linear combination of  $k$  atoms. These parameters play the role of the so-called support of a sparse representation in traditional sparse recovery. In our paper, any kernel and set of parameters that satisfy these conditions are said to be *admissible*.

In the one-dimensional setting, we exhibit a family of kernels relying on *completely monotone functions* for which admissibility holds for any set of atom parameters. For higher dimensional parameter spaces, the analysis turns out to be more subtle. An additional assumption, so-called *axis admissibility*, is imposed to ensure a form of delayed recovery (in at most  $k^D$  steps, where  $D$  is the dimension of the parameter space). Furthermore, guarantees for recovery in exactly  $k$  steps are derived under an additional algebraic condition involving a finite subset of atoms (built as an extension of the set of atoms to be recovered). We show that the latter technical conditions simplify in the case of Laplacian kernels, allowing us to derive simple conditions for  $k$ -step exact recovery, and to carry out a coherence-based analysis in terms of a minimum separation assumption between the atoms to be recovered.

**Keywords:** sparse representation, continuous dictionaries, Orthogonal Matching Pursuit, exact recovery

---

## 1. Introduction

Finding a sparse signal representation is a fundamental problem in signal processing. It consists in decomposing a signal  $\mathbf{y}$  belonging to some vector space  $\mathcal{H}$  as the linear combination of a few elements of some set  $\mathcal{A} \subseteq \mathcal{H}$ , that is

$$\mathbf{y} = \sum_{\ell=1}^k c_{\ell} \mathbf{a}_{\ell} \quad \text{where } c_{\ell} \in \mathbb{R}^*, \mathbf{a}_{\ell} \in \mathcal{A}. \quad (1.1)$$

Sparsity refers to the fact that the number of elements involved in the decomposition (1.1) should be much smaller than the ambient dimension, *i.e.*, the dimension of  $\mathcal{H}$ . The set  $\mathcal{A}$  is commonly referred to as a *dictionary* and its elements as *atoms*. In the sequel, we will assume that  $\mathcal{A}$  is a parametric dictionary defined as:

$$\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \Theta\} \quad (1.2)$$

where  $\Theta = \mathbb{R}^D$  and  $\mathbf{a} : \Theta \rightarrow \mathcal{H}$  is some continuous and injective function. In this setup, (1.1) implies that there exist  $k$  parameters  $\{\theta_{\ell}^*\}_{\ell=1}^k$  such that  $\mathbf{y}$  can be expressed as a linear combination of the atoms  $\{\mathbf{a}(\theta_{\ell}^*)\}_{\ell=1}^k$ .

Over the past decade, sparse representations have proven to be of great interest in many applicative domains. As a consequence, numerous practical procedures, along with their theoretical analyses, have been proposed in the literature. Most contributions addressed the sparse-representation problem in the “*discrete*” setting, where the dictionary contains a *finite* number of elements, see [1]. Recently, several works tackled the problem of sparse representations in “*continuous*” dictionaries, where  $\mathcal{A}$  is made up of an infinite *uncountable* number of atoms but  $\mathbf{a} : \Theta \rightarrow \mathcal{H}$  enjoys some continuity property, see *e.g.*, [2–4]. We review the contributions most related to the present work in Section 2.

Before dwelling over the state of the art, we briefly describe the scope of our paper. In this work, we focus on the continuous setting and assume that  $\mathcal{H}$  is a Hilbert space with inner product  $\langle \cdot, \cdot \rangle$  and induced norm  $\|\cdot\|$ . We derive exact recovery conditions for “Orthogonal Matching Pursuit” (OMP) [5], a natural adaptation to the continuous setting of a popular greedy procedure of the literature (see Algorithm 1). The main question addressed in this paper is as follows. Let  $\{\theta_{\ell}^*\}_{\ell=1}^k$  be  $k$  pairwise distinct elements of  $\Theta$  and assume that  $\mathbf{y}$  obeys (1.1) with  $\mathbf{a}_{\ell} = \mathbf{a}(\theta_{\ell}^*)$  for some  $\{c_{\ell}\}_{\ell=1}^k \subset \mathbb{R}^*$ . Under which conditions does OMP achieve exact recovery (that is, correct unambiguous identification) of the parameters  $\{\theta_{\ell}^*\}_{\ell=1}^k$  and the coefficients  $\{c_{\ell}\}_{\ell=1}^k$ ? In particular, is exact recovery possible in  $k$  steps? This is of course only possible if the preimage  $\theta_{\ell}^*$  of an atom  $\mathbf{a}_{\ell} = \mathbf{a}(\theta_{\ell}^*)$  is unique, hence the assumption that  $\mathbf{a}(\cdot)$  is injective.

We note that, in the context of continuous dictionaries, the fact that OMP could correctly identify a set of  $k$  atoms in exactly  $k$  iterations may seem surprising in itself. Indeed, inspecting Algorithm 1, we see that this implies that OMP must identify one correct atom at *each* iteration  $t$  of the algorithm, that is  $\hat{\theta}_t \in \{\theta_{\ell}^*\}_{\ell=1}^k \forall t \in \llbracket 1, k \rrbracket$ . The following simple example suggests that such a requirement may never be met for continuous dictionaries:

---

**Algorithm 1:** Orthogonal Matching Pursuit (OMP)

---

**Input:** observation  $\mathbf{y} \in \mathcal{H}$ , normalized dictionary  $\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \Theta\}$ .

```

1  $\mathbf{r} \leftarrow \mathbf{y}$  // residual vector
2  $\hat{\mathcal{S}} \leftarrow \emptyset$  // estimated support
3  $t \leftarrow 0$ ;
4 while  $\mathbf{r} \neq \mathbf{0}_{\mathcal{H}}$  do
5    $t \leftarrow t + 1$ ;
6    $\hat{\theta}_t \in \arg \max_{\theta \in \Theta} |\langle \mathbf{a}(\theta), \mathbf{r} \rangle|$  // atom selection
7    $\hat{\mathcal{S}} \leftarrow \hat{\mathcal{S}} \cup \{\hat{\theta}_t\}$  // support update
8    $(\hat{c}_1, \dots, \hat{c}_t) \leftarrow \arg \min_{(c_1, \dots, c_t) \in \mathbb{R}^t} \left\| \mathbf{y} - \sum_{\ell=1}^t c_{\ell} \mathbf{a}(\hat{\theta}_{\ell}) \right\|$  // least-squares update
9    $\mathbf{r} \leftarrow \mathbf{y} - \sum_{\ell=1}^t \hat{c}_{\ell} \mathbf{a}(\hat{\theta}_{\ell})$  // residual vector
10 end
11  $\hat{k} = t$ ;
Output: estimated support  $\hat{\mathcal{S}} = \{\hat{\theta}_{\ell}\}_{\ell=1}^{\hat{k}}$  and coefficients  $\{\hat{c}_{\ell}\}_{\ell=1}^{\hat{k}}$ .

```

---

**Example 1** (The Gaussian deconvolution problem). Consider  $\Theta = \mathbb{R}$  and let  $\mathcal{H} = L^2(\mathbb{R})$  be the space of square integrable functions on  $\mathbb{R}$ . Assume  $\mathbf{a}(\cdot)$  is defined as

$$\begin{aligned} \mathbf{a} : \mathbb{R} &\longrightarrow L^2(\mathbb{R}) \\ \theta &\longmapsto \pi^{-\frac{1}{4}} e^{-\frac{1}{2}(\cdot - \theta)^2}. \end{aligned} \quad (1.3)$$

Suppose  $\mathbf{y}$  results from the positive linear combination of  $k = 2$  distinct atoms, that is  $\mathbf{y} = c_1 \mathbf{a}(\theta_1^*) + c_2 \mathbf{a}(\theta_2^*)$ ,  $\theta_1^* \neq \theta_2^*$ ,  $c_1 > 0$ ,  $c_2 > 0$ . Then, even in this very simple case, OMP never selects an atom in  $\{\theta_1^*, \theta_2^*\}$  at the first iteration. Indeed, particularizing step 6 of Algorithm 1 to the present setup, we have that, at the first iteration, OMP will select the parameter  $\theta$  maximizing

$$|\langle \mathbf{a}(\theta), \mathbf{y} \rangle| = c_1 e^{-\frac{1}{4}(\theta - \theta_1^*)^2} + c_2 e^{-\frac{1}{4}(\theta - \theta_2^*)^2}. \quad (1.4)$$

Now, since the right-hand side of (1.4) is continuously differentiable, first-order optimality conditions tell us that any maximizer of  $\theta \mapsto |\langle \mathbf{a}(\theta), \mathbf{y} \rangle|$  must satisfy

$$(\theta - \theta_1^*) c_1 e^{-\frac{1}{4}(\theta - \theta_1^*)^2} + (\theta - \theta_2^*) c_2 e^{-\frac{1}{4}(\theta - \theta_2^*)^2} = 0. \quad (1.5)$$

Since  $\theta_1^* \neq \theta_2^*$ ,  $c_1 \neq 0$ ,  $c_2 \neq 0$ , this equality cannot be verified by either  $\theta = \theta_1^*$  or  $\theta = \theta_2^*$ . As a consequence, OMP necessarily selects some  $\theta \notin \{\theta_1^*, \theta_2^*\}$ .  $\square$

Nevertheless, we show in this paper that exact recovery in  $k$  steps is possible with OMP for some *particular* families of dictionaries  $\mathcal{A}$ . Our recovery

conditions are expressed in terms of the *kernel function*  $\kappa(\theta, \theta')$  associated to the inner product between two atoms, *i.e.*,

$$\kappa(\theta, \theta') \triangleq \langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle. \quad (1.6)$$

We show that if the kernel  $\kappa(\theta, \theta')$  and the atom parameters  $\{\theta_\ell^*\}_{\ell=1}^k$  verify some particular conditions (see Section 3.2), then exact recovery in  $k$  steps is possible with OMP. We emphasize moreover that these conditions are satisfied for a family of kernels of the form:

$$\kappa(\theta, \theta') = \varphi\left(\|\theta - \theta'\|_p^p\right) \quad 0 < p \leq 1, \quad (1.7)$$

where  $\|\cdot\|_p$  is the  $\ell_p$  quasi-norm (norm for  $p = 1$ ) and  $\varphi$  is a *completely monotone* function (see Definition 3). This family encompasses the well-known Laplace kernel [6]. Hereafter, we will refer to kernels taking the form (1.7) as “CMF kernels”.

A first (perhaps surprising) outcome of our analysis is as follows. If  $\Theta = \mathbb{R}$  and the dictionary is defined by a CMF kernel (1.7), OMP correctly identifies *any* pairwise distinct atom parameters  $\{\theta_\ell^*\}_{\ell=1}^k \subset \Theta$  and coefficients  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$  in exactly  $k$  iterations for *any*  $k \in \mathbb{N}$  (see Theorem 3). We emphasize that no separation (*i.e.*, minimal distance between parameters  $\{\theta_\ell^*\}_{\ell=1}^k$ ) is needed. To our knowledge, this is the first recovery of this kind in continuous dictionaries when no sign constraint is imposed on the coefficients. It turns out that this “universal” exact recovery result is valid for very particular families of dictionaries: CMF kernels exhibit a discontinuity in their derivatives (*e.g.*, the partial derivative of  $\kappa$  with respect to  $\theta$  when  $\theta = \theta'$ ) and the space  $\mathcal{H}$  in which the corresponding dictionary lives is necessarily infinite-dimensional (see Section 3.3).

When  $\Theta = \mathbb{R}^D$  with  $D > 1$  and the dictionary is defined by a CMF kernel (1.7), we show that such an exact recovery result no longer holds (see Example 4). Nevertheless, for dictionaries based on CMF kernels, under an additional hypothesis (referred to as “*axis admissibility*”, see Definition 7), we demonstrate that a form of *delayed* exact recovery (that is, in more than  $k$  iterations) holds. The number of iterations sufficient to identify a set of  $k$  parameters is then upper-bounded by  $k^D$  (see Theorem 4). Moreover, under the above-mentioned hypothesis of axis admissibility, sufficient and necessary conditions for exact recovery of a given subset<sup>1</sup>  $\{\theta_\ell^*\}_{\ell=1}^k$  in  $k$  steps (irrespective of the choice of the coefficients  $\{c_\ell\}_{\ell=1}^k$ ) can be written in terms of a *finite* number of atoms of the dictionary (smaller than  $k^D$ ) including  $\{\theta_\ell^*\}_{\ell=1}^k$  (see Theorem 4). We leverage this result to prove that exact recovery in  $k$  steps is possible as soon as the elements of the subset  $\{\theta_\ell^*\}_{\ell=1}^k$  obey some “minimum separation” condition (see Theorem 5).

The rest of this paper is organized as follows. Section 2 draws connections with the sparse recovery literature. Section 3.1 elaborates on the main ingredients of the “continuous” setup and defines the notions of recovery that are used

<sup>1</sup>Here and in the sequel, when referring to a subset  $\{\theta_\ell^*\}_{\ell=1}^k$ , we implicitly assume that the elements  $\theta_\ell^*$  are pairwise distinct.

in the statements of our results. In Section 3.2, we exhibit a sufficient condition on atom parameters and kernel such that exact recovery of a *given* set of atom parameters holds. We then present the family of CMF dictionaries in Section 3.3 and show in Section 3.4 that different forms of recovery can be achieved in these dictionaries. Concluding remarks are given in Section 4. The technical details of our results are contained in the appendices of the paper. The proofs of our main recovery results are exposed in Appendices A and B. Appendix C contains some auxiliary technical details. Finally, Appendices D and E are dedicated to some mathematical developments related to two examples discussed in the paper.

## Notations

The following notations will be used in this paper. The symbols  $\mathbb{R}, \mathbb{R}^*, \mathbb{R}_+, \mathbb{R}_+^*$  refer to the set of real, non-zero, non-negative and positive numbers, respectively. Boldface lower and upper cases (*e.g.*,  $\mathbf{g}, \mathbf{G}$ ) are used to denote (finite-dimensional) vectors and matrices, respectively. The notation  $[i]$  refers to the  $i$ th element of a vector, and  $[i, j]$  for the element at the  $i$ -th row and  $j$ -th column of a matrix. Italic boldface letters (*e.g.*,  $\mathbf{y}$  or  $\mathbf{a}$ ) denote elements of a Hilbert space  $\mathcal{H}$ . All-one and all-zero column vectors in  $\mathbb{R}^k$  are denoted  $\mathbf{1}_k$  and  $\mathbf{0}_k$ , respectively. The  $\ell$ -th vector of the canonical basis in  $\mathbb{R}^D$  will be denoted  $\mathbf{e}_\ell$ . The notations  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  refer to the inner product and its induced norm on  $\mathcal{H}$ , while  $\|\cdot\|_p$  with  $p > 0$  refers to the classical  $\ell_p$  (pseudo-) norm on  $\mathbb{R}^D$ . Finally, calligraphic letters (*e.g.*,  $\mathcal{S}, \mathcal{G}$ ) are used to describe finite subsets of the parameter space  $\Theta$ , while  $\llbracket m, n \rrbracket$  denotes the set of integers  $i$  such that  $m \leq i \leq n$ . Given  $\mathcal{S} \subseteq \Theta$ , we let  $\mathcal{S}^c \triangleq \Theta \setminus \mathcal{S}$  be the complementary set of  $\mathcal{S}$  in  $\Theta$ . The cardinality of a set is denoted  $\text{card}(\cdot)$ . Finally, if  $\varphi : \mathbb{R} \mapsto \mathbb{R}$  is a function, the notation  $\varphi^{(n)}$  refers to its  $n$ -th derivative. The main notations used in this paper are summarized in Appendix F.

## 2. Related works and state of the art

Over the last decade, sparse representations have sparked a surge of interest in the signal processing, statistics, and machine learning communities. A question of broad interest which has been addressed by many scientists is the identification of the “sparsest” representation of an input signal  $\mathbf{y}$  (that is, the representation involving the smallest number of elements of  $\mathcal{A}$ ). Since this problem has been shown to be NP-hard [7], many sub-optimal procedures have been proposed to approximate its solution.<sup>2</sup> Among the most popular, one can mention methodologies based on *convex relaxation* and *greedy algorithms*.

---

<sup>2</sup>The term “sub-optimal” has to be understood in the following sense: these procedures are heuristics that only find the sparsest solution of the input vector  $\mathbf{y}$  under some restricted conditions. They can fail when these conditions do not hold.

Greedy procedures have a long history in the signal processing and statistical literature, which can be traced back to (at least) the 60's [8]. In the signal processing community, the most popular instances of greedy algorithms are known under the names of *Matching Pursuit* (MP) [9], *Orthogonal Matching Pursuit* (OMP) [5] (also known as *Orthogonal Greedy Algorithm* (OGA) [10, 11]) and *Orthogonal Least Squares* (OLS) [12]. Although these algorithms were already known under different names in other communities [13], they have been “rediscovered” many times, see *e.g.*, [14–16]. Extensions to more general cost functions and kernel dictionaries are discussed in [17].

Sparse representations based on the resolution of convex optimization problems were initially proposed in geophysics [18] for seismic exploration. These methods have been popularized in the signal processing community by the seminal work by Chen *et al.* [19] and by Tibshirani in Statistics [20]. Well-known instances of convex-relaxation approaches for sparse representations are *Basis Pursuit* (BP) [19] and *Lasso* [20], also known as *Basis Pursuit Denoising*, which correspond to different convex optimization formulations. Many algorithmic solutions to efficiently address these problems have been proposed, see *e.g.*, [21–23].

All the early contributions mentioned above have been made in the *discrete* setting, where the dictionary contains a *finite* number of atoms. Although Mallat and Zhang [9] already defined MP for continuous dictionaries, the wide practice of MP is in the discrete setting. Greedy sparse approximation in the context of dictionaries made up of an infinite (possibly uncountable) number of atoms has only been studied more recently [16, 24, 25]. Practical procedures to implement greedy procedures in continuous dictionaries can be found in [26–28].

On the side of convex relaxation approaches, it was shown that a *continuous* version of Lasso can be expressed as a convex optimization problem over the space of Radon measures [29] and later referred to as the *Beurling Lasso* (BLasso) [30]. A continuous version of BP was also proposed [3] for specific continuous dictionaries by exploiting similar ingredients. Motivated by an increasing demand in efficient solvers, different strategies to find the solution of this problem (to some accuracy) were proposed over the past few years. When dealing with dictionaries made up of complex exponentials that depend on a one-dimensional parameter (that is  $D = 1$ ), the BLasso problem can be reformulated as a semidefinite program (SDP) [3, 31]. These methods have been further extended to the multidimensional case by considering SDP approximations of the problem [32]. The conditional gradient method (CGM) has also proven to be applicable to address the BLasso problem [29] and further enhanced with non-convex local optimization extra steps [33–35]. Interestingly, the CGM has been shown to be equivalent to the so-called exchange method in [36, 37]. More recently, gradient-flow methods on spaces of measures have also been investigated to address the BLasso problem [38, 39]. Finally, we also mention the existence of a vast literature on non-convex and non-variational procedures leveraging the celebrated Prony's method [40]. Among others, one may cite its extension to the multivariate case [41], the MUSIC [42] and ESPRIT [43] frameworks, as well as finite rate of innovation methods [44].

Because (most of) the approaches mentioned above (both in the discrete and continuous settings) are heuristics looking for the sparsest representation of some  $\mathbf{y}$ , many theoretical works have been carried out to analyze their performance. Hereafter, we review the contributions of the literature most related to the present work. In particular, we focus on the contributions dealing with exact recovery of some subset  $\{\theta_\ell^*\}_{\ell=1}^k$  for any choice of the coefficients  $\{c_\ell\}_{\ell=1}^k$  (sometimes assuming some specific sign patterns). In our discussion, we will use the short-hand notation  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  and refer to the latter as “*support*”. Since we always implicitly assume that the parameters  $\{\theta_\ell^*\}_{\ell=1}^k$  are pairwise distinct, we have  $\text{card}(\mathcal{S}^*) = k$ . The presentation is organized in two parts, dealing respectively with the discrete and the continuous cases. In the discrete setting, we restrict our attention to contributions addressing the performance of MP, OMP and OLS, *i.e.*, the greedy procedures the most connected to the framework of this paper. In the continuous setting, recovery analysis, including stability and robustness to noise, have only been addressed for convex-relaxation approaches. We review these conditions below and draw some similarities and differences with the guarantees derived for OMP.

### 2.1. Discrete setting

The discrete setting refers to the case where the dictionary contains a finite number of elements, that is  $\text{card}(\mathcal{A}) < \infty$ . Hereafter, we will restrict our discussion to parametric dictionaries of the form (1.2) since they are the main focus of this paper. In this context, the discrete setting refers to  $\text{card}(\Theta) < +\infty$ .

*Exact Recovery Condition.* The first thorough analysis of OMP exact “ $k$ -step” recovery of some  $\mathcal{S}^* \triangleq \{\theta_\ell^*\}_{\ell=1}^k$  is due to Tropp in [45]. Introducing the notations

$$\begin{aligned} \mathbf{G}[\ell, \ell'] &\triangleq \kappa(\theta_\ell^*, \theta_{\ell'}^*), \\ \mathbf{g}_\theta[\ell] &\triangleq \kappa(\theta, \theta_\ell^*), \end{aligned} \quad (2.1)$$

Tropp’s result can be rephrased as follows:

**Theorem 1** (Tropp’s ERC). *Consider  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  and assume that the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$  are linearly independent. If*

$$\forall \theta \in \Theta \setminus \mathcal{S}^*, \quad \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 < 1, \quad (2.2 - \text{ERC})$$

*then OMP with  $\mathbf{y} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  as input unambiguously identifies  $\mathcal{S}^*$  and  $\{c_\ell\}_{\ell=1}^k$  in  $k$  iterations for any choice of the coefficients  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$ . Conversely, if (2.2 – ERC) is not satisfied, there exist not all-zero coefficients  $\{c_\ell\}_{\ell=1}^k$  such that OMP with  $\mathbf{y} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  as input selects some  $\theta \notin \mathcal{S}^*$  at the first iteration.*

A proof of the direct part of this result can be found in [45, Th. 3.1]. The converse part is a slight variation of Tropp’s original statement [45, Th. 3.10] and a proof can be found in *e.g.*, [1, Prop. 3.15].



Condition (2.2 – ERC) is usually referred to as the “Exact Recovery Condition” in the literature, and simply denoted ERC. Assuming linear independence of the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$ , it can be reformulated in the following (and perhaps more interpretable) way:

$$\forall \mathbf{r} \in \mathcal{R}_{\mathcal{S}^*} \setminus \{\mathbf{0}_{\mathcal{H}}\}, \forall \theta \in \Theta \setminus \mathcal{S}^*, \quad |\langle \mathbf{a}(\theta), \mathbf{r} \rangle| < \max_{\theta' \in \mathcal{S}^*} |\langle \mathbf{a}(\theta'), \mathbf{r} \rangle| \quad (2.3)$$

where  $\mathcal{R}_{\mathcal{S}^*} \triangleq \text{span}(\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k)$ . In other words, it implies that OMP *always* selects a parameter in  $\mathcal{S}^*$  during the first  $k$  iterations for any input vector  $\mathbf{y}$  resulting from the linear combination of the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$ . The converse part shows that (2.2 – ERC) is worst-case necessary in the following sense: if (2.2 – ERC) is not satisfied, there exists (at least) one non-trivial linear combination of the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$  such that OMP selects an element  $\theta \notin \mathcal{S}^*$  at the first iteration; in this case the correct identification of  $\mathcal{S}^*$  in  $k$  iterations is obviously not possible.

Interestingly, condition (2.2 – ERC) is also related (along with the linear independence of the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$ ) to the success of MP, OLS and some convex relaxation procedures. In [46, Th. 2], the authors showed that (2.2 – ERC) is also necessary and sufficient for exact  $k$ -step recovery of  $\mathcal{S}^*$  by OLS. Regarding MP, (2.2 – ERC) ensures that the procedure only selects atoms in  $\mathcal{S}^*$  but does not imply exact recovery after  $k$  iterations of the algorithm since the same atom can be selected many times (the least-squares update of the coefficients in Algorithm 1 is not carried out), see *e.g.*, [47, Th. 1]. Finally, in [48, Th. 3] [49, Th. 8], the authors show that (2.2 – ERC) also ensures correct identification of  $\mathcal{S}^*$  by some convex relaxation procedures as *e.g.*, BP or Lasso.

*Coherence.* Tropp’s condition is of limited practical interest to characterize the recovery of all supports of size  $k$  since it requires to verify that (2.2 – ERC) holds for any  $\mathcal{S}^*$  with  $\text{card}(\mathcal{S}^*) = k$ . In order to circumvent this issue, other sufficient conditions of success, weaker but easier to evaluate in practice, have been proposed in the literature. One of the most popular conditions is based on the *coherence*  $\mu$  of a normalized dictionary. Assuming the atoms of the dictionary are of unit norm, this condition writes (with the convention that  $\frac{1}{0} = +\infty$ ):

$$k < \frac{1}{2} \left( 1 + \frac{1}{\mu} \right) \quad (2.4)$$

where

$$\mu \triangleq \sup_{\substack{\theta, \theta' \in \Theta \\ \theta \neq \theta'}} |\kappa(\theta, \theta')|. \quad (2.5)$$

Condition (2.4), together with the normalization of the dictionary, implies that (2.2 – ERC) is verified for any  $\mathcal{S}^*$  with  $\text{card}(\mathcal{S}^*) \leq k$ , and also implies the linear independence of any group of  $k$  atoms of the dictionary. It therefore implies that OMP and OLS correctly identify any  $\mathcal{S}^*$  with  $\text{card}(\mathcal{S}^*) \leq k$  in exactly  $\text{card}(\mathcal{S}^*)$  iterations. It also ensures the correct identification of any  $\mathcal{S}^*$  with  $\text{card}(\mathcal{S}^*) \leq k$

by BP and Lasso. In [50], the authors emphasize that condition (2.4) can be slightly relaxed if the coefficients  $\{c_\ell\}_{\ell=1}^k$  exhibit some decay.

The coherence of the dictionary can be seen as a particular measure of “proximity”<sup>3</sup> between the atoms of the dictionary. Other exact recovery conditions, based on different proximity measures, have been proposed in the literature. In [45, Th. 3.5], the author derived recovery conditions based on “cumulative coherence”, whereas in [11, 51–57], guarantees based on “restricted isometry constants” were proposed. Given that Tropp’s condition is both necessary and sufficient, all such recovery conditions imply that the ERC holds for any support of size  $k$ .

## 2.2. Continuous setting

*General setup.* Sparse representations in continuous dictionaries are basically characterized by two main ingredients:

- i) a parameter set  $\Theta$ , usually assumed to be a connected subset of  $\mathbb{R}^D$  with non-empty interior<sup>4</sup> or a torus in dimension  $D$ . We note that, in this paper, we restrict our attention to the case where  $\Theta = \mathbb{R}^D$  for  $D \geq 1$ .
- ii) an “atom” function  $\mathbf{a} : \Theta \rightarrow \mathcal{H}$ , assumed to be continuous and injective.

This type of dictionary appears in numerous signal processing tasks such as sparse spike deconvolution or super-resolution where one aims to recover fine-scale details from an under-resolved input signal [3, 18, 35, 59].

*Irrelevance of existing analyses.* The continuity of  $\mathbf{a}(\cdot)$  does not allow most of the analyses performed in the context of discrete dictionaries to be extended to the continuous framework. In particular, all exact recovery conditions based on coherence or restricted isometry constants turn out to be violated whenever dealing with continuous dictionaries. As for the coherence condition (2.4), it is easy to see that the continuity of  $\mathbf{a}(\cdot)$  implies the continuity of  $\kappa(\cdot, \cdot)$  with respect to both its arguments. This, in turn, implies that  $\mu = 1$  (for normalized atoms) and the coherence-based condition (2.4) is never met, even for  $k = 1$ !

In order to circumvent this issue, some specific exact recovery conditions for continuous dictionaries have been proposed in the literature, see *e.g.*, [3, 4, 30, 60]. We review below the main ingredients grounding these conditions of recovery. In the context of convex-relaxation approaches, these conditions originate from the analysis of the associated optimality conditions.

*A separation condition for BP for continuous dictionaries.* In the context of BP for continuous dictionaries, the question of exact recovery can be rephrased as follows: if the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$  are linearly independent and  $\mathbf{y} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$ , is the solution of BP for continuous dictionaries unique and equal to a discrete

<sup>3</sup> $\mu = 0$  if all the atoms are pairwise orthogonal and  $\mu \simeq 1$  if some atoms are very correlated.

<sup>4</sup>These two assumptions implies that  $\Theta$  is uncountable [58, Ch. 1, Exercise 19d].

measure supported on  $\{\theta_\ell^*\}_{\ell=1}^k$  with weights  $\{c_\ell\}_{\ell=1}^k$ ?

The case where each atom is a collection of Fourier coefficients ( $\mathcal{H} = \mathbb{C}^m$ ) has received a lot of attention due to its connection with the super-resolution problem. Indeed, the latter scenario is equivalent to recovering infinitely resolved details (the parameters) from some low-pass observation. Without further assumptions on the coefficients, the targeted measure is the unique solution of BP for continuous dictionaries provided that [3, Th. 1.2 and 1.3]

$$\min_{\substack{\ell, \ell' \in \llbracket 1, k \rrbracket \\ \ell \neq \ell'}} |\theta_{\ell'}^* - \theta_\ell^*| > \frac{C}{f_c}, \quad (2.6)$$

where  $|\cdot - \cdot|$  is the  $\ell_\infty$  distance on the  $D$ -dimensional torus (maximum deviation in any coordinate),  $C$  is a constant that depends on the parameter dimension  $D$  and  $f_c$  is the cut-off frequency of the observation low-pass filter. A framing of the value of  $C$  has been proposed by the same authors, further refined in [4, Cor. 1] and [61, Th. 2.2].

We see that (2.6) implies the recovery of  $\{\theta_\ell^*\}_{\ell=1}^k$  and  $\{c_\ell\}_{\ell=1}^k$  provided that the elements of  $\{\theta_\ell^*\}_{\ell=1}^k$  verify some “minimum separation” condition. Interestingly, as shown in [30, Th. 2.1], this separation condition is no longer needed when dealing with *positive* linear combination of atoms (that is, when all the coefficients  $\{c_\ell\}_{\ell=1}^k$  are positive). The authors showed moreover that this separation-free result for positive linear combinations holds for any dictionary such that the atom function  $\mathbf{a}$  forms a “Chebyshev system” [62, Ch. 2] and provided that  $2k + 1$  observations are available (*i.e.*,  $\dim(\mathcal{H}) \geq 2k + 1$ ).

*Dual certificates for the BLasso problem.* In [4], the authors derived several dual certificates for the BLasso problem generalizing the work done by Fuchs for the Lasso [48] to an infinite-dimensional setup. They first show that the existence of a “vanishing derivative pre-certificate” [4, Def. 6] is necessary so that the support of the solution to the BLasso problem is exactly  $\{\theta_\ell^*\}_{\ell=1}^k$  [4, Prop. 8]. On the other hand, they show that a so-called “non degenerate source condition” [4, Def. 5] is sufficient to ensure the desired recovery [4, Th. 2]. We note that these two conditions apply on dictionaries made up of differentiable atom functions  $\mathbf{a} : \Theta \mapsto \mathcal{H}$  since they involve the first and second order derivatives of the inner product  $\langle \mathbf{a}(\theta), \mathbf{y} \rangle$  evaluated at  $\{\theta_\ell^*\}_{\ell=1}^k$ . This is in contrast with the “CMF” dictionaries considered in this paper which involve some non-differentiability in their kernel function, see Section 3.3. Moreover, we also emphasize that both conditions involve the sign of the coefficients  $\{c_\ell\}_{\ell=1}^k$ .

The comparison with the discrete case goes even deeper: it can be shown that the solution of BP for continuous dictionaries is, in some sense, the limit of the solution of BLasso [4, Prop. 1]. Although out of scope of the present paper, we also mention the existence of a literature related to the robustness of BLasso in various noisy settings [4, 63, 64].

### 3. Main results

In this section, we present the main results of the paper. In Section 3.1, we describe the constitutive ingredients of the “continuous” setup addressed in this work and provide a rigorous definition of the notions of exact recovery that will be used in our statements. Our main results are presented in Sections 3.2 and 3.4. The family of “CMF dictionaries”, central to our results in Section 3.4, is introduced in Section 3.3.

#### 3.1. Main ingredients

We first present the three main properties that a “continuous” dictionary should verify, see (3.4a), (3.4b) and (3.6) below. We then elaborate on some differences between the implementation of OMP in the discrete and continuous settings. We finally give a precise definition of the notions of recovery that will be used in the statements of our results.

*Continuous dictionary.* First, the space  $\Theta$  is usually assumed to be a connected metric space or a torus in dimension  $D$ . Hereafter, for the sake of conciseness, we will restrict our attention to the case where  $\Theta = \mathbb{R}^D$  and assume that the kernel associated to the atoms obeys some vanishing property (see (3.6) below).<sup>5</sup> A second common working hypothesis in the “continuous” setup is the continuity of function  $\mathbf{a} : \Theta \rightarrow \mathcal{H}$ , that is

$$\lim_{\theta' \rightarrow \theta} \|\mathbf{a}(\theta') - \mathbf{a}(\theta)\| = 0 \quad \forall \theta \in \Theta. \quad (3.1)$$

In this paper, we will moreover suppose that the atoms of the dictionary are normalized:

$$\|\mathbf{a}(\theta)\| = 1 \quad \forall \theta \in \Theta. \quad (3.2)$$

In the sequel, recovery conditions will be expressed as a function of the induced kernel  $\kappa(\theta, \theta')$ :

$$\kappa(\theta, \theta') \triangleq \langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle \quad \forall \theta, \theta' \in \Theta. \quad (3.3)$$

The “*continuity*” and “*unit-norm*” properties are equivalent to:

$$\text{“unit norm”} : \quad \kappa(\theta, \theta) = 1 \quad \forall \theta \in \Theta, \quad (3.4a)$$

$$\text{“continuity”} : \quad \lim_{\theta' \rightarrow \theta} \kappa(\theta, \theta') = 1 \quad \forall \theta \in \Theta. \quad (3.4b)$$

Moreover, we have from the Cauchy-Schwarz inequality that

$$|\kappa(\theta, \theta')| \leq 1, \quad \forall \theta, \theta' \in \Theta. \quad (3.5)$$

---

<sup>5</sup>Our results can be adapted to any set  $\Theta$  such that step 6 of Algorithm 1 is well-posed, that is, at least one maximizer exists.

Lastly, in this work, we will restrict our attention to kernels that vanish at infinity, *i.e.*,

$$\forall \varepsilon > 0, \forall \theta \in \Theta, \exists K \text{ compact: } \sup_{\theta' \in K^c} \kappa(\theta', \theta) < \varepsilon, \quad (3.6)$$

where  $K^c$  is the complement of  $K$  in  $\Theta$ . This covers the case where  $\Theta$  is compact, by simply considering  $K = \Theta$  with the convention  $\sup_{\theta' \in \emptyset} \kappa = 0$ .

*OMP in continuous dictionaries.* Although Algorithm 1 corresponds to the standard definition of OMP in the discrete setting, its implementation in continuous dictionaries leads to two major differences. First, the “atom selection” step in Line 6 does not necessarily admit a maximizer. In such a case, the recursions defined in Algorithm 1 are ill-posed since the procedure cannot elucidate the maximization problem in Line 6. Second, even if a maximizer exists, solving the “atom selection” problem may be computationally intractable. In particular, the function to be maximized in Line 6 may have many local maxima and the problem is indeed NP-hard in certain cases (*e.g.*, when the maximization step involves a rank-1 approximation of a tensor [65, Th. 1.13] as in [66]), while in other cases it is easy (for example, the SVD can be revisited in this framework and solved up to numerical precision [67, Sec. 2.2]). The maximization problem of Line 6 also appears in Frank-Wolfe type algorithms [35], where existing theoretical guarantees also hold under the hypothesis that this step can be solved.

In this paper, for the sake of simplifying our theoretical analysis, we will nevertheless stick to the idealistic version of OMP described in Algorithm 1. The results presented in this work should therefore be considered more for the theoretical insights they provide into the behavior of OMP in continuous dictionaries than for their practical implications.

In our theoretical analysis we will only have to deal with residuals  $\mathbf{r}$  which can be written as a linear combination of a finite number of atoms of  $\mathcal{A}$ . In such a case, the following lemma shows that a maximizer to the “atom selection” problem always exists:

**Lemma 1.** *Let  $\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \Theta\}$  be a continuous dictionary with kernel  $\kappa(\theta, \theta') \triangleq \langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle$  verifying the continuity property (3.4b) and vanishing property (3.6). Then*

$$\arg \max_{\theta \in \Theta} |\langle \mathbf{a}(\theta), \mathbf{r} \rangle| \neq \emptyset \quad (3.7)$$

*whenever  $\mathbf{r} \in \mathcal{H} \setminus \{\mathbf{0}_{\mathcal{H}}\}$  is a finite linear combination of elements of  $\mathcal{A}$ .*

A proof of this statement is available in Appendix C.1.

We finally emphasize that the solution to the OMP recursions may not be unique. Indeed, in situations where the “atom selection” problem in Line 6 admits *several* solutions, there may exist several output sets,  $\hat{\mathcal{S}}$  and  $\{\hat{c}_\ell\}_{\ell=1}^k$ , verifying OMP recursions. Hereafter, given an observation vector  $\mathbf{y}$ , we will call any set  $\hat{\mathcal{S}}$  which can be generated by OMP with  $\mathbf{y}$  as input, as a “reachable support”.

*Notions of recovery.* The recovery results stated in the next sections of the paper will involve the following notions of success: “*exact  $k$ -step recovery of  $\mathcal{S}^*$* ” and “*exact  $\mathcal{S}$ -delayed recovery of  $\mathcal{S}^*$* ”. We devote the remainder of this section to rigorously defining these two notions.

We say that OMP achieves exact recovery of coefficients  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$  and atom parameters  $\mathcal{S}^* \triangleq \{\theta_\ell^*\}_{\ell=1}^k$  if  $\{c_\ell\}_{\ell=1}^k$  and  $\mathcal{S}^*$  can be unambiguously identified from *any* reachable outputs of OMP ( $\widehat{\mathcal{S}}$  and  $\{\widehat{c}_\ell\}_{\ell=1}^k$ ) run with  $\mathbf{y} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  as input. We note that a simple necessary and sufficient condition for exact recovery of  $\{c_\ell\}_{\ell=1}^k$  and  $\mathcal{S}^*$  reads

$$\mathcal{S}^* \subseteq \widehat{\mathcal{S}}, \quad (3.8)$$

for each reachable support  $\widehat{\mathcal{S}}$ . This can be seen from the following arguments. If there is a reachable support such that  $\widehat{\mathcal{S}} \not\supseteq \mathcal{S}^*$ , then exact recovery is obviously not attained since there exists some  $\theta_\ell^* \in \mathcal{S}^*$  that is not identified in  $\widehat{\mathcal{S}}$ . Conversely, if  $\mathcal{S}^* \subseteq \widehat{\mathcal{S}}$  holds, one must have

$$\forall \ell \in [1, \widehat{k}], \begin{cases} \widehat{c}_\ell = c_\ell & \text{if } \theta_\ell^* \in \mathcal{S}^* \\ \widehat{c}_\ell = 0 & \text{otherwise,} \end{cases} \quad (3.9)$$

because the atoms  $\{\mathbf{a}(\theta) : \theta \in \widehat{\mathcal{S}}\}$  selected by OMP are always linearly independent and  $\mathbf{y} \in \text{span}(\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k)$ . Therefore,  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  can be unambiguously identified from the non-zero elements of  $\{\widehat{c}_\ell\}_{\ell=1}^k$ .

In the literature related to the conditions of success of OMP, a distinction is usually made between the cases “ $\mathcal{S}^* = \widehat{\mathcal{S}}$ ” and “ $\mathcal{S}^* \subseteq \widehat{\mathcal{S}}$ ”: the former is referred to as “ *$k$ -step recovery*” because it implies that OMP identifies  $\mathcal{S}^*$  and  $\{c_\ell\}_{\ell=1}^k$  in exactly  $k$  steps; the latter is known as “*delayed recovery*” because OMP may require (if the inclusion is strict) to carry out more than  $k$  iterations to identify  $\mathcal{S}^*$  and  $\{c_\ell\}_{\ell=1}^k$ .

In this paper we will focus on conditions ensuring the correct identification of a given support  $\mathcal{S}^*$  of cardinality  $k$  for *any* choice of the non-zero weighting coefficients  $\{c_\ell\}_{\ell=1}^k$ . The notions of “*exact  $k$ -step recovery of  $\mathcal{S}^*$* ” and “*exact  $\mathcal{S}$ -delayed recovery of  $\mathcal{S}^*$* ” announced at the beginning of this section then read as follows. We say that OMP achieves “*exact  $k$ -step recovery of  $\mathcal{S}^*$* ” if  $\mathcal{S}^* = \widehat{\mathcal{S}}$  for any choice of  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$  and any reachable output  $\widehat{\mathcal{S}}$ . This implies that there is only one reachable output. Moreover, given some set  $\mathcal{S} \subset \Theta$ , we say that OMP achieves “*exact  $\mathcal{S}$ -delayed recovery of  $\mathcal{S}^*$* ” if

$$\mathcal{S}^* \subseteq \widehat{\mathcal{S}} \subseteq \mathcal{S} \quad (3.10)$$

for any choice of  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$  and any reachable output  $\widehat{\mathcal{S}}$ . “ *$\mathcal{S}$ -delayed recovery*” can be regarded as a refined version of “*delayed recovery*” where the set of parameters that OMP may select is guaranteed to belong to some set  $\mathcal{S} \supseteq \mathcal{S}^*$ . Uninterestingly this is always the case with  $\mathcal{S} = \Theta$ , so what will be important in our results is to establish conditions such that we can identify a *finite* set  $\mathcal{S}$ ,

determined by the only specification of  $\mathcal{S}^*$ , such that  $\mathcal{S}$ -delayed recovery of  $\mathcal{S}^*$  holds. We note that  $\mathcal{S}$ -delayed recovery implies that OMP identifies  $\mathcal{S}^*$  in at most  $\text{card}(\mathcal{S})$  iterations. Finally, we emphasize that “exact  $\mathcal{S}^*$ -delayed recovery of  $\mathcal{S}^*$ ” is equivalent to “exact  $k$ -step recovery of  $\mathcal{S}^*$ ”. We will sometimes use the former in the formulation of our results to have more compact statements. We will also always implicitly assume that OMP achieves  $k$ -step recovery of  $\mathcal{S}^*$  when  $\mathcal{S}^* = \emptyset$  since this implies that  $\mathbf{y} = \mathbf{0}_{\mathcal{H}}$  and OMP returns the empty support  $\hat{\mathcal{S}} = \emptyset$  at iteration 0 in this case.

### 3.2. Exact recovery of a given support: sufficient conditions

In this section, we highlight some instrumental properties of the dictionary  $\mathcal{A}$  and support  $\mathcal{S}^*$  which allow OMP to achieve exact  $\text{card}(\mathcal{S})$ -step recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$  (see Theorem 2). These conditions are the basis of our results on “CMF dictionaries” stated in Section 3.4.

We first notice that, in the context of continuous dictionaries, the  $k$ -step analysis of Theorem 1 still applies: condition (2.2 – ERC) along with the linear independence of the atoms  $\{\mathbf{a}(\theta_\ell)\}_{\ell=1}^k$  are still necessary and sufficient for exact recovery of a support  $\mathcal{S}^*$ .<sup>6</sup> However, the standard formulation

$$\max_{\theta \in \Theta \setminus \mathcal{S}^*} \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 < 1, \quad (3.11)$$

equivalent to (2.2 – ERC) in the discrete setting, does no longer hold in the case of continuous dictionaries as the supremum

$$\sup_{\theta \in \Theta \setminus \mathcal{S}^*} \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 \quad (3.12)$$

is always at least 1.<sup>7</sup>

In order to circumvent this problem, we identify below two simpler conditions, respectively on the dictionary  $\mathcal{A}$  (via its induced kernel  $\kappa$ ) and the support  $\mathcal{S}^*$ , which imply that the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$  are linearly independent and that (2.2 – ERC) is verified, see Theorem 2 below. The following definition includes assumptions on the kernel ensuring that the dictionary atoms are normalized, and that the atom function  $\theta \mapsto \mathbf{a}(\theta)$  is injective and continuous.

**Definition 1** (Admissible kernel). *A kernel  $\kappa$  is said to be admissible if:*

- i) *it verifies (3.4) and (3.6).*
- ii)  *$0 \leq \kappa(\theta, \theta') < 1$  for any  $\theta \neq \theta'$ .*

*By extension, a dictionary  $\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \Theta\}$  is said to be admissible if its induced kernel is admissible.*

<sup>6</sup>We note in particular that (2.2 – ERC) ensures that the “atom selection” step in Line 6 of Algorithm 1 is well-defined since the maximizers are ensured to belong to the finite set  $\mathcal{S}^*$ .

<sup>7</sup>Indeed, first notice that  $\|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 = 1$  for all  $\theta \in \mathcal{S}^*$ . One then obtains that the supremum is at least 1 by continuity of  $\theta \mapsto \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1$ .

**Definition 2** (Admissible support with respect to kernel  $\kappa$ ). A support  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  is admissible with respect to a kernel  $\kappa$  if the following holds for any non-empty subset  $T \subseteq \llbracket 1, k \rrbracket$  and any positive coefficients  $\{c_\ell\}_{\ell \in T} \subset \mathbb{R}_+^*$  such that  $\sum_{\ell \in T} c_\ell < 1$ :

i) The set of global maximizers of

$$\begin{aligned} \psi: \Theta &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \sum_{\ell \in T} c_\ell \kappa(\theta, \theta_\ell^*), \end{aligned} \quad (3.13)$$

is a subset of  $\{\theta_\ell^*\}_{\ell \in T}$ .

ii) If  $\ell \in \llbracket 1, k \rrbracket \setminus T$  satisfies  $\psi(\theta) - \kappa(\theta, \theta_\ell^*) \leq 0$  for all  $\theta \in \{\theta_{\ell'}^*\}_{\ell' \in T}$ , then

$$\forall \theta \in \Theta, \quad \psi(\theta) - \kappa(\theta, \theta_\ell^*) \leq 0. \quad (3.14)$$

By extension, the support  $\mathcal{S}^*$  is said to be admissible with respect to dictionary  $\mathcal{A} = \{\mathbf{a}(\theta): \theta \in \Theta\}$  if  $\mathcal{S}^*$  is admissible with respect to the kernel induced by  $\mathcal{A}$ .

With these definitions, our first recovery result reads:

**Theorem 2.** Assume  $\mathcal{A}$  is admissible and  $\mathcal{S}^*$  is admissible with respect to  $\mathcal{A}$ . Then, OMP achieves exact  $\text{card}(\mathcal{S})$ -step recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$ .

A proof of this result is available in Appendix A. Theorem 2 provides some sufficient conditions for exact  $\text{card}(\mathcal{S})$ -step recovery of any  $\mathcal{S} \subseteq \mathcal{S}^*$  via the definitions of “admissible dictionary” and “admissible support”. In particular, the conditions of Theorem 2 imply that (2.2 – ERC) is satisfied for any  $\mathcal{S} \subseteq \mathcal{S}^*$ . As we will see in Section 3.4, the admissibility of  $\mathcal{A}$  and  $\mathcal{S}^*$  may be much easier to prove in some cases than verifying directly that (2.2 – ERC) holds.

As the admissibility conditions stated in Definitions 1 and 2 may appear somewhat technical, we discuss hereafter the different items appearing in these definitions in order to shed some light on the scope of Theorem 2.

In Definition 1, (3.4) ensures that the kernel  $\kappa$  induced by  $\mathcal{A}$  is continuous and that the dictionary  $\mathcal{A}$  only contains unit-norm atoms. The continuity assumption is crucial in the derivation of our result since it induces a specific structure on the dictionary. The unit-norm hypothesis is only secondary but allows to avoid some unnecessary technicalities in the proofs. Hypothesis (3.6) ensures the well-posedness of the “atom selection” step in Line 6 of Algorithm 1, see Lemma 1. Finally,  $0 \leq \kappa(\theta, \theta')$  implies that the inner product between two atoms of  $\mathcal{A}$  is always nonnegative, whereas  $\kappa(\theta, \theta') < 1$  guarantees that  $\mathbf{a}(\theta) \neq \mathbf{a}(\theta')$  for  $\theta \neq \theta'$ , i.e., that  $\theta \mapsto \mathbf{a}(\theta)$  is an injective function (remember that we assume  $\kappa(\theta, \theta) = 1$  for all  $\theta \in \Theta$ ; the fact that atoms are distinct is thus a direct consequence of the Cauchy-Schwarz inequality). The atoms  $\mathbf{a}(\theta) \neq \mathbf{a}(\theta')$  for  $\theta \neq \theta'$  being normalized, distinct, and positively correlated, they are also linearly independent.



As for Definition 2, item *i)* ensures that a correct atom selection always occurs when the residual  $\mathbf{r}$  is a *positive* combination<sup>8</sup> of the atoms of the support and the kernel  $\kappa$  is admissible. Indeed, if  $\mathbf{r} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  with  $c_1, \dots, c_k > 0$  and the kernel is admissible then from Definition 1

$$|\langle \mathbf{a}(\theta), \mathbf{r} \rangle| = \sum_{\ell=1}^k c_\ell \kappa(\theta, \theta_\ell^*). \quad (3.15)$$

In such a case, item *i)* of Definition 2 then implies

$$\arg \max_{\theta \in \Theta} |\langle \mathbf{a}(\theta), \mathbf{r} \rangle| \subseteq \mathcal{S}^*. \quad (3.16)$$

Item *ii)* of Definition 2 does not have such a simple interpretation but a careful inspection of our proof in Appendix A shows that this condition is instrumental for deriving the result stated in Theorem 2. Altogether, given some admissible dictionary  $\mathcal{A}$ , Theorem 2 allows us to establish recovery results valid without sign constraints by only proving the two assumptions gathered in Definition 2, which somehow correspond to establishing the result for the easier case of positive combinations of atoms.

### 3.3. CMF dictionaries

In the next section, we will particularize Theorem 2 to a family of dictionaries whose kernel is defined via a completely monotone function (CMF). In this section, we provide a precise definition of this family of dictionaries and some of their properties that will be used throughout the paper.

We first recall the definition of a CMF:

**Definition 3** (CMF [68, Def. 7.1]). *A function  $\varphi : \mathbb{R}_+ \mapsto \mathbb{R}$  is completely monotone on  $[0, +\infty[$  if it is infinitely differentiable on  $]0, +\infty[$ , right continuous at 0, and if its derivatives obey*

$$(-1)^n \varphi^{(n)}(x) \geq 0 \quad \forall x, n \in \mathbb{R}_+^* \times \mathbb{N}. \quad (3.17)$$

As described in the following example, many well-known functions are CMFs:

**Example 2.** *The following functions are completely monotone [6]:*

- *the function  $x \mapsto e^{-\lambda x}$  for  $\lambda > 0$  which gives birth to the Laplace kernel,*
- *the function  $x \mapsto \frac{1}{1+\lambda x}$  for  $\lambda > 0$ ,*
- *ratios of modified Bessel functions of the first kind,*
- *a subset of the confluent hypergeometric functions (Kummer's function),*

---

<sup>8</sup> In particular, we note that if item *i)* of Definition 2 is true, then its conclusion still holds without the hypothesis “ $\sum_{\ell \in T} c_\ell < 1$ ”.

- a subset of the Gauss hypergeometric functions.

By definition, CMFs are non-negative, non-increasing and convex functions. Moreover, they admit an integral formulation in terms of Laplace transform of a Borel measure:

**Lemma 2** (Bernstein-Widder theorem, [68, Th. 7.11]). *A function  $\varphi$  is completely monotone on  $[0, +\infty[$  if and only if there exists a non-negative finite measure  $\nu$  on Borel sets of  $[0, +\infty[$  such that*

$$\varphi(x) = \int_0^{+\infty} e^{-ux} d\nu(u), \quad (3.18)$$

where the integral converges for all  $x \geq 0$  since  $\nu$  is finite and  $e^{-ux} \leq 1$ .

We note for example that the Laplace kernel (see Example 2) is a CMF with representation measure equal to  $\nu = \delta_\lambda$  with  $\lambda > 0$ .

In the sequel we will consider the following class of kernels and dictionaries whose definitions rely on the concept of CMF:

**Definition 4** (CMF kernel and dictionary). *The class of CMF kernels in dimension  $D \geq 1$ , denoted  $\mathcal{K}_{\text{CMF}}(D)$ , consists of all kernels  $\kappa : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}_+^*$  such that*

$$\kappa(\theta, \theta') = \varphi\left(\|\theta - \theta'\|_p^p\right) \quad \forall \theta, \theta' \in \mathbb{R}^D \quad (3.19)$$

where  $\varphi$  is a CMF verifying  $\varphi(0) = 1$ ,  $\lim_{x \rightarrow +\infty} \varphi(x) = 0$  and  $0 < p \leq 1$ .

By extension, we say that  $\mathcal{A}$  is a CMF dictionary in dimension  $D \geq 1$  if its induced kernel belongs to  $\mathcal{K}_{\text{CMF}}(D)$ .

We note that the constraint  $\varphi(0) = 1$  in the previous definition ensures that the “unit-norm” hypothesis (3.4a) is satisfied. We also mention that the constraint  $\lim_{x \rightarrow +\infty} \varphi(x) = 0$  is necessary so that CMF kernels satisfy the vanishing property (3.6).

At this point, a legitimate question is whether kernel (3.19) can be induced by some dictionary  $\mathcal{A}$ ? The answer is positive and is a corollary of the following lemma:

**Lemma 3.** *Let  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$  be a CMF such that  $\varphi(0) = 1$ ,  $\lim_{x \rightarrow \infty} \varphi(x) = 0$  and  $0 < p \leq 1$ . Then, any function of the form*

$$\begin{aligned} \rho : \mathbb{R}^D &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \varphi(\|\omega\|_p^p) \end{aligned} \quad (3.20)$$

is positive definite.

A proof of this result is provided in Appendix B.1. We refer the reader to [68, Def. 6.1] for a precise definition of positive (semi-) definite functions and [68, Th. 6.2] for a review of some of their basic properties. In particular, the positive definite nature of  $\varphi(\|\cdot\|_p^p)$  used in conjunction with standard results in the theory of “reproducing kernel Hilbert spaces” (see *e.g.*, [69, Th. 3.11]) implies the following corollary:

**Corollary 1** (Existence of CMF dictionaries). *For any  $\kappa \in \mathcal{K}_{\text{CMF}}(D)$ , there exists some Hilbert space  $\mathcal{H}$  and some (continuous) function  $\mathbf{a} : \mathbb{R}^D \rightarrow \mathcal{H}$  such that (3.3) holds. Moreover, any finite collection of distinct elements from  $\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \mathbb{R}^D\}$  is linearly independent.*

We see from the last part of the corollary that CMF dictionaries are necessarily defined in infinite-dimensional Hilbert spaces  $\mathcal{H}$ . If not, any collection of  $\dim(\mathcal{H}) + 1$  elements of  $\mathcal{A}$  would be linearly dependent which is in contradiction with Corollary 1. The next example exhibits a family of atoms in  $\mathcal{H} = L^2(\mathbb{R})$  which is a CMF dictionary in  $\mathbb{R}$ .

**Example 3.** *Let  $\Theta = \mathbb{R}$  and consider the dictionary  $\mathcal{A}$  defined by*

$$\begin{aligned} \mathbf{a} : \mathbb{R} &\longrightarrow L^2(\mathbb{R}) \\ \theta &\longmapsto f_\theta(t) = \sqrt{2\lambda} e^{-\lambda(t-\theta)} \mathbf{1}_{\{t \geq \theta\}}(t) \end{aligned} \quad (3.21)$$

for some  $\lambda > 0$ , where  $\mathbf{1}_{\{t \geq \theta\}}$  is the “indicator” function which is equal to 1 if  $t \geq \theta$  and 0 otherwise. Straightforward calculations both show that  $\|\mathbf{a}(\theta)\| = 1$  for any  $\theta$  and the inner product in  $\mathcal{H} = L^2(\mathbb{R})$  between two atoms writes  $\langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle = e^{-\lambda|\theta - \theta'|}$ . The latter function corresponds to the so-called “Laplace kernel”. This kernel is an element of  $\mathcal{K}_{\text{CMF}}(1)$  according to Example 2.

We conclude this section by introducing a particular CMF kernel which will be used in the statement of some of our results in Section 3.4:

**Definition 5** (Generalized Laplace kernel and dictionary). *The class of Generalized Laplace kernels in dimension  $D$ , denoted  $\mathcal{K}_{\text{Lap}}(D)$ , consists of all kernels  $\kappa : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}_+^*$  such that*

$$\kappa(\theta, \theta') = e^{-\lambda \|\theta - \theta'\|_p^p} \quad \forall \theta, \theta' \in \mathbb{R}^D \quad (3.22)$$

where  $\lambda > 0$  and  $0 < p \leq 1$ .

By extension, a Generalized Laplace dictionary in dimension  $D \geq 1$  is a collection of atoms  $\mathcal{A} = \{\mathbf{a}(\theta) : \theta \in \mathbb{R}^D\}$  whose induced kernel belongs to  $\mathcal{K}_{\text{Lap}}(D)$ .

One immediately sees that  $\mathcal{K}_{\text{Lap}}(D) \subset \mathcal{K}_{\text{CMF}}(D)$  since the function  $t \mapsto e^{-\lambda t}$  defined on  $\mathbb{R}_+$  is a CMF (see Example 2).

### 3.4. Recovery conditions in CMF dictionaries

In this section, we provide recovery results for OMP in CMF dictionaries. The proofs of our results are based on the sufficient conditions presented in Theorem 2 and are reported in Appendix B. A first surprising result holds when  $\Theta = \mathbb{R}$ :

**Theorem 3.** *Assume  $\mathcal{A}$  is a CMF dictionary in dimension 1. Then, OMP achieves exact  $\text{card}(\mathcal{S}^*)$ -step recovery of each finite support  $\mathcal{S}^* \subset \mathbb{R}$ .*

In essence, Theorem 3 identifies a class of dictionaries for which exact  $k$ -step recovery is possible for *any* support  $\mathcal{S}^*$  of *any* finite size  $k$ . We note that the notions of exact recovery of a support  $\mathcal{S}^*$  defined in Section 3.1 do not involve any sign constraint on the coefficients  $\{c_\ell\}_{\ell=1}^k$  used to generate the observation vector  $\mathbf{y}$ . As a comparison, the results ensuring the success of continuous BP/BLasso with no sign constraints on  $\{c_\ell\}_{\ell=1}^k$  require some “minimum separation condition” between parameters  $\{\theta_\ell^*\}_{\ell=1}^k$  to hold (see (2.6) and related discussion). Conversely, the recovery results for BLasso obtained in [30] without separation condition require weighting coefficients  $\{c_\ell\}_{\ell=1}^k$  to be positive. The novelty of Theorem 3 is thus a separation-free recovery result for any signed finite linear combination of atoms. The strength of the result obtained in Theorem 3 comes however at a price: it applies to a specific family of dictionaries, namely CMF dictionaries. In particular, as mentioned in Section 3.3, the space  $\mathcal{H}$  in which CMF dictionaries live is necessarily infinite-dimensional, and the corresponding kernels exhibit a discontinuity in all their partial derivatives at  $\theta = \theta' \in \Theta$ . Another price to pay is that the recovery guarantees are for OMP, an algorithm explicitly involving the search for the global maximum of an optimization problem, cf Line 6 of Algorithm 1.

In higher dimension  $D > 1$ , the “universal” exact recovery result stated in Theorem 3 no longer holds, as shown in the next example. More precisely, if  $D \geq 3$ , we emphasize that there always exists a configuration of parameters  $\{\theta_\ell^*\}_{\ell=1}^k$  such that OMP fails at the first iteration for some  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$ :

**Example 4.** Let  $D \geq 3$  and  $3 \leq k \leq D$ . Consider  $\mathcal{S}^* \triangleq \{\theta_\ell^*\}_{\ell=1}^k \subset \mathbb{R}^D$  and  $\Delta > 0$  such that

$$\begin{aligned} \|\theta_\ell^* - \theta_{\ell'}^*\|_p^p &= 2\Delta^p \quad \forall \ell \neq \ell' \\ \|\theta_\ell^* - \mathbf{0}_D\|_p^p &= \Delta^p \quad \forall \ell. \end{aligned}$$

Let  $\mathbf{a} : \mathbb{R}^D \mapsto \mathcal{H}$  define a CMF dictionary in  $\mathbb{R}^D$  with kernel  $\kappa = \varphi(\|\cdot - \cdot\|_p^p)$ . We next show that, if  $\Delta$  is sufficiently small, there always exists a linear combination of  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$  such that OMP selects a parameter not in  $\mathcal{S}^*$  at the first iteration.

Let us consider  $\mathbf{y} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  and assume that all coefficients  $c_\ell$  are equal. We then have

$$\frac{\langle \mathbf{a}(\mathbf{0}_D), \mathbf{y} \rangle}{\langle \mathbf{a}(\theta_\ell^*), \mathbf{y} \rangle} = \frac{k\varphi(\Delta^p)}{1 + (k-1)\varphi(2\Delta^p)}. \quad (3.23)$$

Then,  $\theta = \mathbf{0}_D$  will be preferred to all “ground-truth” parameters  $\theta_\ell^*$  at the first iteration of OMP as soon as the quantity in (3.23) is larger than 1, or, equivalently,

$$(k-1)\varphi(2\Delta^p) - k\varphi(\Delta^p) + 1 < 0. \quad (3.24)$$

Let us show that (3.24) holds whenever  $\Delta^p$  is “sufficiently small”. For simplicity, consider first the case where  $\varphi(t) = e^{-\lambda t}$  with  $\lambda > 0$ . Condition (3.24) writes

$$(k-1)x^2 - kx + 1 < 0 \quad (3.25)$$

with  $x = \varphi(\Delta^p) = e^{-\lambda\Delta^p}$ . As  $k \geq 3$ , the left-hand side of (3.25) is a second order polynomial with two distinct roots, namely  $(k-1)^{-1}$  and 1. Therefore, OMP prefers  $\mathbf{0}_D$  as soon as  $(k-1)^{-1} < x < 1$  or, equivalently,  $\Delta^p < \lambda^{-1} \log(k-1)$ . The latter condition implies a necessary separation condition such that OMP does not fail at the first iteration. We note that it is possible to draw similar conclusions whenever  $\varphi$  is a CMF function right-differentiable at zero. The proof of this result requires extra work that is detailed in Appendix D.

Although a “universal”  $k$ -step recovery result such as Theorem 3 no longer holds in CMF dictionaries when  $D > 1$ , it is nevertheless possible to show that some form of exact recovery of a support  $\mathcal{S}^*$  is possible under an additional condition on the kernel induced by the CMF dictionary (see Theorem 4). This additional condition is referred to as “axis admissibility” hereafter and is encapsulated in Definition 7 below. Before moving on to this definition, it is necessary to introduce the notions of “Cartesian grid” and “set augementer operator”:

**Definition 6** (Cartesian grid). A finite set  $\mathcal{G} \subset \mathbb{R}^D$  is a Cartesian grid in dimension  $D \geq 1$  if there exists  $D$  one-dimensional finite sets  $\{\mathcal{S}_d\}_{d=1}^D$  such that

$$\mathcal{G} = \prod_{d=1}^D \mathcal{S}_d, \quad (3.26)$$

where  $\prod$  denotes the Cartesian product.

We moreover define the following “set augementer” operator that, given a finite set  $\mathcal{S} \subset \mathbb{R}^D$ , returns the smallest Cartesian grid containing  $\mathcal{S}$ :

$$\text{Grid}(\mathcal{S}) \triangleq \prod_{d=1}^D \left\{ \theta[d] : \theta \in \mathcal{S} \right\}. \quad (3.27)$$

It is quite straightforward to see that  $\mathcal{S} \subseteq \text{Grid}(\mathcal{S})$  for any finite set  $\mathcal{S} \subset \mathbb{R}^D$  and that the operator  $\text{Grid}$  is idempotent. We illustrate the definition of  $\text{Grid}(\mathcal{S})$  in Fig. 1 in dimension  $D = 2$  for  $\mathcal{S} = \{\theta_1, \theta_2, \theta_3\}$ .

We are now ready to introduce the notion of “axis admissibility”:

**Definition 7** (Axis admissibility with respect to a kernel). A Cartesian grid  $\mathcal{G} = \prod_{d=1}^D \mathcal{S}_d = \{\theta_\ell\}_{\ell=1}^{\text{card}(\mathcal{G})}$  is said to be axis admissible with respect to a kernel  $\kappa$  if and only if  $\forall d \in \llbracket 1, D \rrbracket$ ,  $\forall \theta \in \mathbb{R}^D$  with  $\theta[d] = 0$  and  $\forall \{c_\ell\}_{\ell=1}^{\text{card}(\mathcal{G})} \subset \mathbb{R}$  such that the function

$$f_d(t) = \left| \sum_{\ell=1}^{\text{card}(\mathcal{G})} c_\ell \kappa(\theta + t\mathbf{e}_d, \theta_\ell) \right| \quad (3.28)$$

is not identically zero, we have

$$\emptyset \neq \arg \max_{t \in \mathbb{R}} f_d(t) \subseteq \mathcal{S}_d. \quad (3.29)$$

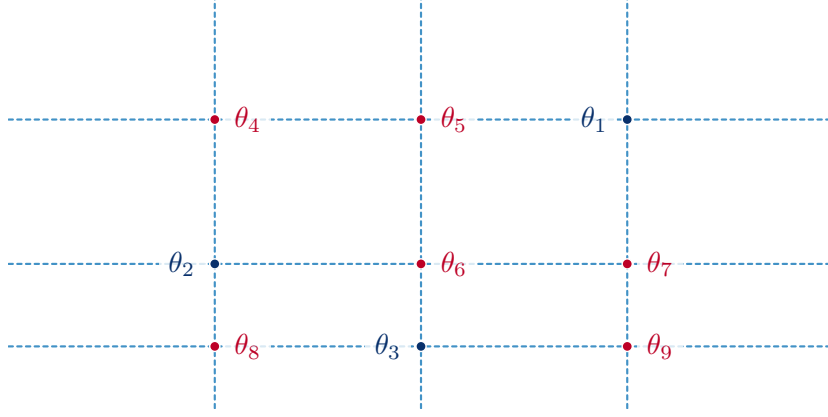


Figure 1: Illustration in dimension  $D = 2$  with  $k = 3$  of the definition of the set augmenter  $\text{Grid}$  defined in (3.27). The blue points, denoted  $\theta_\ell$  for  $\ell \in \{1, 2, 3\}$ , form the support  $\mathcal{S}$ . The red points, denoted  $\theta_\ell$ ,  $\ell \in \llbracket 4, 9 \rrbracket$  represent the elements of  $\text{Grid}(\mathcal{S}) \setminus \mathcal{S}$ .

By extension, a Cartesian grid  $\mathcal{G}$  is said to be axis admissible with respect to a dictionary  $\mathcal{A}$  if it is axis admissible with respect to the kernel induced by  $\mathcal{A}$ .

The notion of axis admissibility will be central in our next result to ensure the exact recovery of some support  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  in a CMF dictionary. In particular, we will see that axis admissibility of  $\text{Grid}(\mathcal{S}^*)$  ensures exact  $\text{Grid}(\mathcal{S}^*)$ -delayed recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$ . Moreover, exact  $\mathcal{S}^*$ -delayed recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$  is achievable by combining axis admissibility of  $\text{Grid}(\mathcal{S}^*)$  with the following restricted version of the ERC:<sup>9</sup>

$$\max_{\theta \in \text{Grid}(\mathcal{S}^*) \setminus \mathcal{S}^*} \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 < 1 \quad (3.30\text{-R-ERC})$$

where

$$\begin{aligned} \mathbf{G}[\ell, \ell'] &\triangleq \langle \mathbf{a}(\theta_\ell^*), \mathbf{a}(\theta_{\ell'}^*) \rangle \quad \forall \ell, \ell' \in \llbracket 1, k \rrbracket \\ \mathbf{g}_\theta[\ell] &\triangleq \langle \mathbf{a}(\theta), \mathbf{a}(\theta_\ell^*) \rangle \quad \forall \ell \in \llbracket 1, k \rrbracket \end{aligned} \quad (3.31)$$

Formally, our next result writes as follows:

**Theorem 4.** *Let  $\mathcal{A}$  be a CMF dictionary in  $\mathbb{R}^D$  with induced kernel  $\kappa$  and let  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$ .*

- *If  $\text{Grid}(\mathcal{S}^*)$  is axis admissible with respect to  $\kappa$ , then OMP achieves  $\text{Grid}(\mathcal{S}^*)$ -delayed recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$ . If (3.30-R-ERC) moreover holds, OMP achieves  $\mathcal{S}^*$ -delayed recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$ .*

<sup>9</sup>We remind the reader that  $\mathbf{G}$  is invertible as the Gram matrix of a set of linearly independent atoms (see Corollary 1).

- Conversely, if (3.30-R-ERC) does not hold, there exists not all-zero coefficients  $\{c_\ell\}_{\ell=1}^k$  such that OMP with  $\mathbf{y} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  as input selects some  $\theta \notin \mathcal{S}^*$  at the first iteration.

A first outcome of Theorem 4 is a (pessimistic) upper bound on the number of iterations needed to identify  $\mathcal{S}^*$  when  $\mathcal{A}$  is a CMF dictionary and  $\text{Grid}(\mathcal{S}^*)$  is axis admissible with respect to the kernel induced by  $\mathcal{A}$ . In particular, the first part of the theorem states that OMP needs no more than  $\text{card}(\text{Grid}(\mathcal{S}^*)) \leq k^D$  iterations to succeed. As shown in the second part of the theorem, this (rather pessimistic) upper bound on the number of iterations can be decreased to  $k$  if an additional restricted ERC (3.30-R-ERC) is verified. Interestingly, whereas the parameter space  $\Theta$  is a continuum, (3.30-R-ERC) only depends on a finite subset of the elements of  $\Theta$  (namely  $\text{Grid}(\mathcal{S}^*)$ ) and its numerical evaluation is therefore possible. We will see in Theorem 5 below, that this restricted ERC allows us to derive a separability condition for exact  $k$ -step recovery in Generalized Laplace dictionaries. Besides, we note that additional strategies could be investigated to improve the upper bound, exploiting, *e.g.*, coefficients decay [50].

In our next result, we show that the property of “axis admissibility” can be (at least) satisfied for some CMF dictionaries. In particular, the next lemma emphasizes that any Cartesian grid is axis admissible for Generalized Laplace dictionaries (see Definition 5):

**Lemma 4.** *Let  $\mathcal{A}$  be a Generalized Laplace dictionary in  $\mathbb{R}^D$ . Then all Cartesian grids  $\mathcal{G}$  are admissible with respect to  $\mathcal{A}$ .*

A proof of this result is given in Appendix B.4. Combining this lemma with Theorem 4 immediately leads to the following corollary:

**Corollary 2.** *Let  $\mathcal{A}$  be a Generalized Laplace dictionary in  $\mathbb{R}^D$ . Then OMP achieves exact  $\text{Grid}(\mathcal{S}^*)$ -delayed recovery of each finite support  $\mathcal{S}^* \subset \mathbb{R}^D$ .*

Interestingly, although Example 4 showed that exact  $\text{card}(\mathcal{S}^*)$ -step recovery does not hold for arbitrary  $\mathcal{S}^*$  in CMF dictionaries, Corollary 2 emphasizes that exact  $\text{Grid}(\mathcal{S}^*)$ -delayed recovery is achievable by OMP in Generalized Laplace dictionaries for any  $\mathcal{S}^*$  and any  $k = \text{card}(\mathcal{S}^*) \in \mathbb{N}^*$ . Following our remark below Theorem 4, OMP is thus ensured to identify any support of size  $k$  in at most  $k^D$  iterations in this type of dictionaries. Similar to Theorem 3, no separation assumptions nor sign constraints are needed here to ensure our recovery result, although it applies to a very specific family of dictionaries. We will see in Theorem 5 below that adding some separation condition on the elements of  $\mathcal{S}^*$  enables to verify (3.30-R-ERC) and therefore leads to an exact-recovery result in at most  $k$  steps.

Before moving on to the statement of this result, let us mention that, although Lemma 4 shows that any Cartesian grid is axis admissible with respect to Generalized Laplace dictionaries, such a result does in general not hold for CMF dictionaries without extra assumptions on the grid. Nevertheless, our empirical evidence suggests that the admissible grid assumption is only an artifact of our proof technique. We conjecture that Theorem 4 remains valid even

when the Cartesian grid  $\mathcal{G}$  is not axis admissible. To support our conjecture, we show in Appendix E that the second part of Theorem 4 still holds for *any* CMF dictionary and for any  $\mathcal{S}^* \subset \mathbb{R}^D$  with  $\text{card}(\mathcal{S}^*) = 2$ , even though  $\text{Grid}(\mathcal{S}^*)$  is generally not axis admissible. The proof of this kind of result in the general case is still under investigation.

In the last result of this section, we particularize (3.30-R-ERC) to derive a separation condition on the elements of  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  that ensures exact  $\text{card}(\mathcal{S})$ -step recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$  in Generalized Laplace dictionaries. We first note that, following standard results of the literature (see *e.g.*, [45]), (3.30-R-ERC) can be relaxed to a mutual coherence condition:

$$\mu < \frac{1}{2k-1} \quad (3.32)$$

where

$$\mu \triangleq \max_{\substack{\theta, \theta' \in \text{Grid}(\mathcal{S}^*) \\ \theta \neq \theta'}} |\langle \mathbf{a}(\theta), \mathbf{a}(\theta') \rangle|. \quad (3.33)$$

Our separation result is then a simple consequence of this mutual coherence condition:

**Theorem 5.** *Let  $\mathcal{A}$  be a Generalized Laplace dictionary in  $\mathbb{R}^D$  with parameters  $\lambda > 0$  and  $0 < p \leq 1$ . Consider  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  and let*

$$\Delta_0 \triangleq \min_{d \in \llbracket 1, D \rrbracket} \min \{ |\theta_\ell^*[d] - \theta_{\ell'}^*[d]| : \ell, \ell' \in \llbracket 1, k \rrbracket \text{ s.t. } \theta_\ell^*[d] \neq \theta_{\ell'}^*[d] \}. \quad (3.34)$$

If

$$\Delta_0^p \geq \frac{\log(2k-1)}{\lambda} \quad (3.35)$$

then, OMP achieves exact  $\text{card}(\mathcal{S})$ -steps recovery of each  $\mathcal{S} \subseteq \mathcal{S}^*$ .

*Proof.* By definition of  $\Delta_0$  and of  $\text{Grid}(\mathcal{S}^*)$ , we have  $\|\theta - \theta'\|_p^p \geq \Delta_0^p$  for all  $\theta, \theta' \in \text{Grid}(\mathcal{S}^*)$ . Hence, using the definition of the mutual coherence in (3.33) we have  $\mu = \exp(-\lambda \|\theta - \theta'\|_p^p)$  for some  $\theta, \theta' \in \mathcal{S}^*$  so  $\mu \leq \exp(-\lambda \Delta_0^p)$  and (3.35) implies that  $\mu < (2k-1)^{-1}$  holds.  $\square$

Theorem 5 states that, with Generalized Laplace dictionaries, OMP recovers any linear combination of  $k$  sufficiently separated atoms in  $k$  steps. Although condition (3.35) is expressed in terms of minimal distance between parameters, it can be seen as a condition on the mutual coherence between atoms. However, in contrast to the discrete case, this mutual coherence guarantee is only related to a particular finite subset of the (continuous) Generalized Laplace dictionary, namely the atoms with parameters in  $\text{Grid}(\mathcal{S}^*)$ .

Furthermore, condition (3.35) is reminiscent of the separation condition for off-the-grid super-resolution proposed in [3], see (2.6). The so-called *separation condition* discussed in (2.6) is expressed on a  $D$ -dimensional torus preventing also high values of  $k$ . For example, in a unit-length 1-dimensional torus and



with the notations of (2.6), the minimum separation condition for BP requires  $k \leq \frac{f_c}{C} - 1$ . Note however that these results involve different dictionaries and settings making relevant comparison tedious.

#### 4. Conclusion - discussion

In this work, we have shown that the study of the recovery properties of greedy procedures such *Orthogonal Matching Pursuit* (OMP) can be extended to the setting of continuous dictionaries where the atoms continuously depend on some parameters. Capitalizing on the formulation of OMP in terms of inner products between atoms, our results rely on the properties of the kernel implicitly defined by the inner product between atoms. More particularly, we have identified two key notions which we have called *admissible kernel* and *admissible support*, that are sufficient to ensure exact recovery irrespective of the value of the coefficients involved in the representation. For the class of CMF dictionaries, we have shown that when the dimension of the parameter space is 1, all implicitly defined kernels as well as all supports are admissible. Up to our knowledge, this is the first class of kernels for which no separation is needed to achieve exact recovery, even for signed combinations of atoms. However, such a “universal” recovery result comes at a price since CMF dictionaries can only live in infinite-dimensional observation spaces  $\mathcal{H}$  and the corresponding kernels exhibit some discontinuities in their derivatives.

Although exact recovery can also be ensured for CMF dictionaries with a parameter space of dimension greater than 1, extra conditions have to be imposed on the support to be recovered, as some supports may not be admissible anymore. The cornerstone of our analysis in the multi-dimensional case is the notion of *axis admissible* Cartesian grid. Indeed, axis admissibility is sufficient to allow OMP to identify supports, leading to a form of “delayed recovery” for all supports of size  $k$  embedded in some admissible Cartesian grid. For such supports, exact  $k$ -step recovery can also be achieved whenever a condition on a finite number of (known) atoms is fulfilled. In the special case of Generalized Laplace dictionaries, any Cartesian grid turns out to be axis admissible, and a simplified *coherence-based analysis* can be revisited, leading to exact  $k$ -step recovery under a *minimal separation condition*.

We now review some prospects of this work:

*Beyond axis-admissible grids for CMF kernels.* Our analysis for multi-dimensional parameter sets relies on the notion of axis-admissible grids. While axis admissibility holds for any grid with respect to Generalized Laplace dictionaries, this is apparently no longer the case with respect to more general CMF dictionaries. Even for grids which seem to violate the axis-admissibility condition with respect to a CMF dictionary, empirical evidence suggests that Theorem 4 remains valid. As a first step towards a better understanding of this phenomenon, we showed in Appendix E that, for supports of size 2, axis-admissibility is not necessary for the conclusion of Theorem 4 to hold.

*Connection with TV-minimization.* In light of the existing links between Tropp's ERC [45] and recovery guarantees for  $\ell_1$  minimization [49], an interesting question is whether the guarantees developed in this paper can be extended to sparse spike recovery with total variation norm minimization (see Section 2). More particularly, one could benefit from the *null-space properties for measures* [30] which characterize the solution of the continuous version of Basis Pursuit. Such a connection may yield support recovery results for signed combinations of atoms with TV-norm minimization without separation conditions.

*Robustness to estimation error.* In the discrete setting, one advantage of greedy procedures over convex relaxations is that the associated recovery guarantees involve solutions provided by actual *algorithms* rather than merely expressed as the minimizer of some *optimization problem*. In the continuous setting, this has to be tempered with the fact that implementing OMP requires a (possibly intractable) global maximization procedure at each iteration. Our current analysis does not take into account the resulting numerical estimation error or the fact that there may be spurious local maxima. One could envision overcoming some of these limitations by analyzing the behavior of OMP when a small error is systematically done when maximizing the inner product in Line 6 of Algorithm 1. Note that such an approximation error may also be useful to account for discretized implementations of the latter step of OMP using a fine grid over the parameter set  $\Theta$ .

## A. Proof of Theorem 2

Let  $\mathcal{S} \subseteq \mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$ . Without loss of generality, we assume that  $\mathcal{S} \neq \emptyset$  corresponds to the first  $\text{card}(\mathcal{S})$  elements of  $\mathcal{S}^*$ , that is  $\mathcal{S} = \{\theta_\ell^*\}_{\ell=1}^{\text{card}(\mathcal{S})}$ .

We first notice that, as a direct consequence of Definition 2, if  $\mathcal{S}^*$  is admissible with respect to  $\kappa$  then any  $\mathcal{S} \subseteq \mathcal{S}^*$  is also admissible with respect to  $\kappa$ . The result stated in Theorem 2 is then a direct consequence of Theorem 1 and the following proposition:

**Proposition 1.** *Assume kernel  $\kappa$  is admissible and  $\mathcal{S} = \{\theta_\ell^*\}_{\ell=1}^{\text{card}(\mathcal{S})}$  is admissible with respect to  $\kappa$ . Then we have that*

i) *the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k$  are linearly independent,*

ii)  *$\forall \theta \in \Theta \setminus \mathcal{S}, \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 < 1$ , where*

$$\mathbf{G}[\ell, \ell'] \triangleq \kappa(\theta_\ell^*, \theta_{\ell'}^*) \quad \forall \ell, \ell' \in \llbracket 1, \text{card}(\mathcal{S}) \rrbracket \quad (\text{A.1})$$

$$\mathbf{g}_\theta[\ell] \triangleq \kappa(\theta, \theta_\ell^*) \quad \forall \ell \in \llbracket 1, \text{card}(\mathcal{S}) \rrbracket. \quad (\text{A.2})$$

We thus spend the rest of this section in proving Proposition 1.

*Proof of item i) of Proposition 1.* Let  $\{c_\ell\}_{\ell=1}^{\text{card}(\mathcal{S})} \subset \mathbb{R}$  be such that  $\mathbf{y} \triangleq \sum_{\ell=1}^{\text{card}(\mathcal{S})} c_\ell \mathbf{a}(\theta_\ell^*) = \mathbf{0}_{\mathcal{H}}$ , and let  $T$  be the set of indices such that  $c_\ell \neq 0$ . Without loss of generality, we can assume that  $\sum_{\ell=1}^{\text{card}(\mathcal{S})} |c_\ell| < 1$ . We will prove by contradiction that  $T$  is empty.

Assuming that  $T$  is not empty, we first prove that the sign of the coefficients  $\{c_\ell\}_{\ell \in T}$  cannot be all equal. To this end, let us assume (without loss of generality) that  $c_\ell > 0$  for all  $\ell \in T$  and show that a contradiction occurs with the hypothesis of admissibility of  $\mathcal{S}$ . Since  $\mathbf{y} = \mathbf{0}_{\mathcal{H}}$ , the function  $\psi : \theta \mapsto \langle \mathbf{a}(\theta), \mathbf{y} \rangle$  is identically equal to zero. Hence, on the one hand, any point of  $\Theta$  is a maximizer. On the other hand, since all the elements of  $\{c_\ell\}_{\ell \in T}$  are positive and  $\mathcal{S}$  is (by hypothesis) admissible with respect to  $\kappa$ , we have from item i) of Definition 2 that the maximizers of  $\psi$  must belong to  $\mathcal{S}$ . This implies that  $\Theta \subset \mathcal{S}$  which contradicts the definition of  $\mathcal{S}$  and  $\Theta$ . Therefore, if  $T$  is not empty, not all the elements of  $\{c_\ell\}_{\ell \in T}$  have the same sign.

We can thus partition  $T$  into two non-empty disjoint subsets:

$$\begin{aligned} T_+ &= \{\ell \in T : c_\ell > 0\}, \\ T_- &= \{\ell \in T : c_\ell < 0\}. \end{aligned}$$

Similarly, we let

$$\begin{aligned} \mathcal{S}_+ &= \{\theta_\ell^* \in \mathcal{S} : \ell \in T_+\}, \\ \mathcal{S}_- &= \{\theta_\ell^* \in \mathcal{S} : \ell \in T_-\}. \end{aligned}$$

Since the elements of  $\mathcal{S} \supseteq \mathcal{S}_- \cup \mathcal{S}_+$  are pairwise distinct, we have  $\mathcal{S}_- \cap \mathcal{S}_+ = \emptyset$ . Defining  $\mathbf{y}_+ = \sum_{\ell \in T_+} c_\ell \mathbf{a}(\theta_\ell^*)$  and  $\mathbf{y}_- = \sum_{\ell \in T_-} c_\ell \mathbf{a}(\theta_\ell^*)$ , we note that  $\mathbf{y} = \mathbf{y}_+ + \mathbf{y}_-$ . Using the fact that  $\mathbf{y} = \mathbf{0}_{\mathcal{H}}$ , one deduces that  $\mathbf{y}_+ = -\mathbf{y}_-$ . Moreover,  $\mathbf{y}_+$  (resp.  $-\mathbf{y}_-$ ) is a positive linear combination of atoms with parameters in  $\mathcal{S}_+ \subset \mathcal{S}$  (resp.  $\mathcal{S}_- \subset \mathcal{S}$ ). Therefore, since  $\mathcal{S}$  is admissible with respect to  $\kappa$ , item i) of Definition 2 applies and we have that any maximizer of  $\psi : \theta \mapsto \langle \mathbf{a}(\theta), \mathbf{y}_+ \rangle = \langle \mathbf{a}(\theta), -\mathbf{y}_- \rangle$  must belong to  $\mathcal{S}_+ \cap \mathcal{S}_-$ . Now, on the one hand, by Lemma 1, the set of maximizers of  $\psi$  cannot be empty. On the other hand  $\mathcal{S}_+ \cap \mathcal{S}_- = \emptyset$ . This leads to a contradiction. Therefore we must have  $T = \emptyset$ . In other words,  $\mathbf{y} = \mathbf{0}_{\mathcal{H}}$  implies  $c_\ell = 0 \ \forall \ell \in \mathcal{S}$ , so that the atoms  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^{\text{card}(\mathcal{S})}$  are linearly independent.

As a consequence of this first part of the proposition, the Gram matrix of any subset of  $\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^{\text{card}(\mathcal{S})}$  is a positive definite matrix, and therefore invertible. In particular, the inverse of matrices  $\mathbf{G}$  and  $\overline{\mathbf{G}}$  appearing in the second part of the proof is always well-defined.

*Proof of item ii) of Proposition 1.* Recall that, as a consequence of Definition 2, if  $\mathcal{S}$  is admissible with respect to  $\kappa$ , then any support  $\mathcal{S}' \subset \mathcal{S}$  is also admissible. We thus show our result by induction on the cardinality of  $\mathcal{S}'$ . For notational convenience, we let hereafter  $k' \triangleq \text{card}(\mathcal{S}')$ . We prove by induction on  $k'$  that:

a)  $\mathbf{G}^{-1} \mathbf{1}_{k'}$  has nonnegative entries,

b)  $\forall \theta \in \Theta$ ,  $\mathbf{G}^{-1}\mathbf{g}_\theta$  has nonnegative entries,

c)  $\forall \theta \in \Theta \setminus \mathcal{S}'$ ,  $\|\mathbf{G}^{-1}\mathbf{g}_\theta\|_1 < 1$ .

The quantities  $\mathbf{G}$  and  $\mathbf{g}_\theta$  appearing above are defined in (A.1)-(A.2) with the substitution  $\mathcal{S}' \leftrightarrow \mathcal{S}$ . Item c) corresponds to result ii) of Proposition 1. Items a) and b) are intermediate results that allow a subdivision of the proof into steps.

*Initialization:*  $k' = 1$ . In this case, both  $\mathbf{G}$  and  $\mathbf{g}_\theta$  are scalars. Since  $\kappa$  is admissible, we have  $\mathbf{G} = 1$  and  $\mathbf{g}_\theta \geq 0$  (cf Definition 1). Therefore, items a) and b) are fulfilled and  $\|\mathbf{G}^{-1}\mathbf{g}_\theta\|_1 = \mathbf{g}_\theta = \kappa(\theta, \theta_1^*)$  and, using Definition 1-ii), we have  $\kappa(\theta, \theta_1^*) < 1$ . Hence, item c) is also true.

*Induction:*  $1 < k' \leq k$ . We assume items a)-b)-c) hold for any  $\mathcal{S}' \subset \mathcal{S}$  of cardinality  $k' - 1 \geq 1$ . Considering  $\mathcal{S}' \subseteq \mathcal{S}$  an arbitrary support of size  $k'$ , we show that items a)-b)-c) also hold for  $\mathcal{S}'$ . Without loss of generality, we will assume that  $\mathcal{S}'$  corresponds to the first  $\text{card}(\mathcal{S}')$  elements of  $\mathcal{S}$ , that is  $\mathcal{S}' = \{\theta_\ell^*\}_{\ell=1}^{k'}$ .

We consider  $\overline{\mathcal{S}} = \{\theta_\ell^*\}_{\ell=1}^{k'-1} \subset \mathcal{S}'$  and use over-lined notations for quantities related to  $\overline{\mathcal{S}}$ : we denote by  $\overline{\mathbf{G}} \in \mathbb{R}^{(k'-1) \times (k'-1)}$ ,  $\overline{\mathbf{g}}_\theta \in \mathbb{R}^{k'-1}$  the quantities defined in (A.1)-(A.2) for  $\overline{\mathcal{S}}$  and by  $\mathbf{G} \in \mathbb{R}^{k' \times k'}$ ,  $\mathbf{g}_\theta \in \mathbb{R}^{k'}$  the same quantities for  $\mathcal{S}'$ . Likewise, the notations  $\overline{\mathbf{g}}_\ell \in \mathbb{R}^{k'-1}$ ,  $\mathbf{g}_{\ell'} \in \mathbb{R}^{k'}$  for  $\ell = 1 \dots k' - 1$ ,  $\ell' = 1 \dots k'$  will refer to the columns of  $\overline{\mathbf{G}}$  and  $\mathbf{G}$ , respectively. With these notations we have:

$$\mathbf{g}_\theta = \begin{pmatrix} \overline{\mathbf{g}}_\theta \\ \kappa(\theta, \theta_{k'}^*) \end{pmatrix} \in \mathbb{R}^{k'} \quad \forall \theta \in \Theta \quad (\text{A.3})$$

$$\mathbf{G} = \begin{pmatrix} \overline{\mathbf{G}} & \overline{\mathbf{g}}_{k'} \\ \overline{\mathbf{g}}_{k'}^T & 1 \end{pmatrix} \in \mathbb{R}^{k' \times k'} \quad (\text{A.4})$$

where we denote  $\overline{\mathbf{g}}_{k'} \triangleq \overline{\mathbf{g}}_{\theta_{k'}^*}$  for notational convenience. We note that, as mentioned above, item i) of Proposition 1 ensures that both  $\mathbf{G}$  and  $\overline{\mathbf{G}}$  are invertible.

*Item a).* We show that the last entry of  $\mathbf{u} \triangleq \mathbf{G}^{-1}\mathbf{1}_{k'}$  is positive. Since the reasoning holds for any ordering of the  $\theta_\ell^*$ 's, we then deduce that all the entries of  $\mathbf{u}$  are positive. Block inversion results [70, Cor. 2.8.9] give

$$\mathbf{G}^{-1} = \begin{pmatrix} \overline{\mathbf{G}}^{-1} + s\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_{k'}\overline{\mathbf{g}}_{k'}^T\overline{\mathbf{G}}^{-1} & -s\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_{k'} \\ -s\overline{\mathbf{g}}_{k'}^T\overline{\mathbf{G}}^{-1} & s \end{pmatrix}, \quad (\text{A.5})$$

where  $s \triangleq (1 - \overline{\mathbf{g}}_{k'}^T\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_{k'})^{-1}$ . Notice that

$$\begin{aligned} \overline{\mathbf{g}}_{k'}^T\overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_{k'} &\leq \|\overline{\mathbf{g}}_{k'}\|_\infty \left\| \overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_{k'} \right\|_1 \\ &\leq \left\| \overline{\mathbf{G}}^{-1}\overline{\mathbf{g}}_{k'} \right\|_1 \\ &< 1. \end{aligned} \quad (\text{A.6})$$

The first inequality is a consequence of Hölder's inequality, the second of Definition 1 and the third follows from induction hypothesis *c*). Hence  $s > 0$ .

The last entry of  $\mathbf{u} = \mathbf{G}^{-1}\mathbf{1}_{k'}$  now writes  $\mathbf{u}[k'] = s(1 - \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{G}}^{-1} \mathbf{1}_{k'-1})$ . By induction hypothesis *b*), we have  $\|\bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_{k'}\|_1 = \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{G}}^{-1} \mathbf{1}_{k'-1}$ . Using (A.6) and the fact that  $s > 0$ , we thus have  $\mathbf{u}[k'] > 0$ .

*Item b*). We first show that the last entry of  $\mathbf{v} \triangleq \mathbf{G}^{-1} \mathbf{g}_\theta$  is non-negative. Given the decomposition of  $\mathbf{G}$  in (A.5), the last entry of  $\mathbf{G}^{-1} \mathbf{g}_\theta$  writes

$$\mathbf{v}[k'] = s \left( \mathbf{g}_\theta[k'] - \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_\theta \right) = s \left( \kappa(\theta, \theta_{k'}^*) - \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_\theta \right). \quad (\text{A.7})$$

Since  $s > 0$  (see (A.6)) it is then sufficient to show that  $\kappa(\theta, \theta_{k'}^*) - \bar{\mathbf{v}}^T \bar{\mathbf{g}}_\theta \geq 0$ , where  $\bar{\mathbf{v}} \triangleq \bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_{k'}$ , in order to show that  $\mathbf{v}[k'] \geq 0$ . This will be achieved by studying this quantity seen as a function of  $\theta$ . Consider  $T \subseteq \llbracket 1, k' - 1 \rrbracket$  the (possibly empty) set defined by  $T \triangleq \{\ell : \bar{\mathbf{v}}[\ell] \neq 0\}$  and define

$$\begin{aligned} \psi_1 : \Theta &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \bar{\mathbf{v}}^T \bar{\mathbf{g}}_\theta = \sum_{\ell=1}^{k'-1} \bar{\mathbf{v}}[\ell] \kappa(\theta, \theta_\ell^*) = \sum_{\ell \in T} \bar{\mathbf{v}}[\ell] \kappa(\theta, \theta_\ell^*). \end{aligned} \quad (\text{A.8})$$

Notice that:

- $\psi_1(\theta) = \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_\theta$ ,
- the entries of  $\bar{\mathbf{v}}$  are nonnegative by the induction hypothesis *b*). Moreover, from induction hypothesis *c*), we have  $\sum_{\ell=1}^{k'-1} \bar{\mathbf{v}}[\ell] = \|\bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_{k'}\|_1 < 1$ ,
- for  $j \in \llbracket 1, k' - 1 \rrbracket$  and  $\theta = \theta_j^*$  we have  $\bar{\mathbf{g}}_\theta = \bar{\mathbf{g}}_j = \bar{\mathbf{G}} \bar{\mathbf{e}}_j$ , where  $\bar{\mathbf{e}}_j$  is the  $j$ -th canonical vector of  $\mathbb{R}^{k'-1}$ . Hence,  $\psi_1(\theta_j^*) = \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{G}}^{-1} \bar{\mathbf{g}}_j = \bar{\mathbf{g}}_{k'}^T \bar{\mathbf{e}}_j = \kappa(\theta_j^*, \theta_{k'}^*)$  and  $\psi_1(\theta_j^*) - \kappa(\theta_j^*, \theta_{k'}^*) = 0 \forall j \neq k'$ .

Since  $\mathcal{S}$  is admissible with respect to  $\kappa$ , we can apply item *ii*) of Definition 2 with  $\ell = k'$  to any  $\emptyset \neq T \subseteq \llbracket 1, k' - 1 \rrbracket$ . This leads to:

$$\kappa(\theta, \theta_{k'}^*) - \bar{\mathbf{v}}^T \bar{\mathbf{g}}_\theta = \kappa(\theta, \theta_{k'}^*) - \psi_1(\theta) \geq 0 \quad (\text{A.9})$$

for all  $\theta \in \Theta$ . The same obviously holds if  $T$  is empty as  $\psi_1(\theta)$  is identically zero and the admissibility of  $\kappa$  implies that it is nonnegative (see Definition 1). Since this result does not depend on the ordering of the  $\theta_\ell^*$ 's, we can finally conclude that all the elements of  $\mathbf{G}^{-1} \mathbf{g}_\theta$  are nonnegative.

*Item c*). Let

$$\begin{aligned} \psi_2 : \Theta &\longrightarrow \mathbb{R} \\ \theta &\longmapsto \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1. \end{aligned} \quad (\text{A.10})$$

We need to prove that  $\psi_2(\theta) < 1$  for all  $\theta \notin \mathcal{S}'$ .

From item *b)*, we know that  $\mathbf{G}^{-1}\mathbf{g}_\theta$  has nonnegative entries, so that  $\|\mathbf{G}^{-1}\mathbf{g}_\theta\|_1 = \mathbf{1}_{k'}^T \mathbf{G}^{-1}\mathbf{g}_\theta$ . Letting  $\mathbf{u} \triangleq \mathbf{G}^{-1}\mathbf{1}_{k'} \in \mathbb{R}^{k'}$ , the function  $\psi_2(\theta)$  can then be written as

$$\psi_2(\theta) \triangleq \sum_{\ell=1}^{k'} \mathbf{u}[\ell] \kappa(\theta, \theta_\ell^*). \quad (\text{A.11})$$

Moreover we have  $\mathbf{u}[\ell] \geq 0 \forall \ell$  since we showed in item *a)* that  $\mathbf{G}^{-1}\mathbf{1}_{k'} = \mathbf{u}$  has nonnegative entries. We also note that  $\mathbf{u} \neq \mathbf{0}_{k'}$  since  $\mathbf{G}\mathbf{u} = \mathbf{1}_{k'}$ .

Applying item *i)* of Definition 2 together with the comment in Footnote 8, we have that the maximizer of  $\psi_2(\theta)$  must belong to  $\{\theta_\ell^* : \mathbf{u}[\ell] \neq 0\} \subseteq \mathcal{S}'$ . Now,

$$\forall j \in \llbracket 1, k' \rrbracket, \quad \psi_2(\theta_j^*) = \mathbf{1}_{k'}^T \mathbf{G}^{-1} \mathbf{g}_{\theta_j^*} = \mathbf{1}_{k'}^T \mathbf{e}_j = 1. \quad (\text{A.12})$$

Therefore,  $\psi_2(\theta) < 1$  for all  $\theta \notin \mathcal{S}'$ .

## B. Proofs related to CMF dictionaries

This appendix contains the proofs of the results related to CMF dictionaries presented in Sections 3.3 and 3.4. We first state and prove a technical lemma which will be used in the proofs of Lemma 3 and Theorem 3:

**Lemma 5.** *Let  $\varphi$  be a CMF such that  $\varphi(0) = 1$  and  $\lim_{t \rightarrow \infty} \varphi(t) = 0$ . Then the Borel measure  $\nu$  appearing in the integral representation of the CMF in Lemma 2 is nonzero and satisfies  $\nu(\{0\}) = 0$  and  $\nu(\mathbb{R}_+^*) = 1$ . Moreover,  $\varphi$  is strictly positive and strictly decreasing on  $\mathbb{R}_+$ , with  $\varphi^{(1)}(t) < 0$  on  $\mathbb{R}_+$ .*

*Proof.* By the integral representation of Lemma 2 we have  $\nu(\mathbb{R}_+) = \varphi(0) = 1$  hence  $\nu$  is nonzero. Moreover  $\varphi(x) = \int_0^{+\infty} e^{-ux} d\nu(u) \geq \nu(\{0\}) \geq 0$  for each  $x \geq 0$ . As  $\lim_{x \rightarrow +\infty} \varphi(x) = 0$ , it follows that  $\nu(\{0\}) = 0$  and therefore  $\nu(\mathbb{R}_+^*) = 1$ . This proves the first part of the statement.

The positivity of  $\varphi$  follows from the fact that  $\nu$  is nonzero and  $e^{-ux} > 0$  for all  $u, x \geq 0$ . Hence, the integral representation (3.18) of  $\varphi$  yields  $\varphi(x) > 0$  for each  $x > 0$ . Finally, we prove by contradiction that  $\varphi^{(1)}(t) < 0$  on  $\mathbb{R}_+$ .

Assume the existence of  $t_0 > 0$  such that  $\varphi^{(1)}(t_0) = 0$ . As  $\varphi^{(1)}$  is continuous and non-decreasing on  $\mathbb{R}_+^*$  with  $\varphi^{(1)}(t) \leq 0$  for each  $t > 0$ , it follows that  $\varphi^{(1)}(t) = 0$  for each  $t \geq t_0$ , hence  $\varphi(t) = \varphi(t_0)$  for each  $t \geq t_0$ . As we have just seen, we have  $\varphi(t_0) > 0$ , hence this contradicts the assumption  $\lim_{t \rightarrow +\infty} \varphi(t) = 0$ .  $\square$

### B.1. Proof of Lemma 3

The outline of the proof is as follows. We first show that for any  $0 < p \leq 1$  and  $\boldsymbol{\omega} \in \mathbb{R}^D$ , the quantity  $e^{-u\|\boldsymbol{\omega}\|_p^p}$  is related to the characteristic function of some  $D$ -dimensional random vector  $\mathbf{Z}_u$ . We then use this formulation of  $e^{-u\|\boldsymbol{\omega}\|_p^p}$  as a characteristic function together with the Bernstein-Widder representation of CMFs (see Lemma 2) to show that the function  $\varphi(\|\boldsymbol{\omega}\|_p^p)$  is positive definite.

*Proof that  $e^{-u\|\omega\|_p^p}$  is the characteristic function of some random vector  $\mathbf{Z}_u$ .* In probability theory, the characteristic function of a real-valued random vector  $\mathbf{Z} \in \mathbb{R}^D$  is the function  $\omega \in \mathbb{R}^D \mapsto \mathbb{E}_{\mathbf{Z}}[e^{i\omega^T \mathbf{Z}}]$  where  $\mathbb{E}$  denotes the expectation operator and  $i$  is the imaginary number.

First we consider for any  $u \geq 0$  the scalar-valued function

$$\begin{aligned} \Phi_u : \mathbb{R} &\longrightarrow \mathbb{R} \\ \omega &\longmapsto e^{-u|\omega|^p} \end{aligned}$$

and show that for  $u > 0$  it is the characteristic function of some random variable  $Z_u$  which admits a density with respect to the Lebesgue measure. Our proof leverages a result due to Pólya [71, Th. 1]. We reproduce this result hereafter for self-containedness of the paper:

**Theorem 6.** *Let  $\Phi$  be a real-valued function defined on  $\mathbb{R}$  and such that:*

- $\Phi$  is continuous and even,
- $\Phi$  is convex on  $\mathbb{R}_+^*$ ,
- $\Phi(0) = 1$ ,
- $\lim_{\omega \rightarrow +\infty} \Phi(\omega) = 0$ .

*Then,  $\Phi$  is the characteristic function of some random variable which admits a density with respect to the Lebesgue measure. Moreover this density is even and continuous everywhere, except possibly at zero.*

Observe that  $\Phi_u$  is even, continuous and verifies  $\Phi_u(0) = 1$  and  $\lim_{\omega \rightarrow +\infty} \Phi_u(\omega) = 0$  since  $u > 0$ . Moreover, for  $p \in ]0, 1]$  and  $\omega > 0$ , its second derivative on  $\mathbb{R}_+$  is

$$\Phi_u^{(2)}(\omega) = up\omega^{p-2}((1-p) + up\omega^p)e^{-u\omega^p}.$$

Hence,  $\Phi_u^{(2)}(\omega) > 0$  for all  $\omega > 0$  and  $\Phi_u$  is convex on  $\mathbb{R}_+^*$ . As a consequence,  $\Phi_u$  satisfies the assumptions of Theorem 6 and it is the characteristic function of some scalar random variable  $Z_u$  which admits a (continuous, except possibly at zero) density with respect to the Lebesgue measure, that is

$$\Phi_u(\omega) = \mathbb{E}_{Z_u}[e^{i\omega Z_u}].$$

We are now ready to show that the function  $e^{-u\|\omega\|_p^p}$  is the characteristic function of some random vector  $\mathbf{Z}_u$ . To this end, let us define the random vector  $\mathbf{Z}_u = (Z_u^1, \dots, Z_u^D)^T$  as the concatenation of  $D$  independent copies of  $Z_u$ . We thus have

$$\mathbb{E}_{\mathbf{Z}_u}[e^{i\omega^T \mathbf{Z}_u}] = \prod_{d=1}^D \mathbb{E}_{Z_u^d}[e^{i\omega[d] Z_u^d}] = \prod_{d=1}^D \Phi_u(\omega[d]) = e^{-u\|\omega\|_p^p}.$$

We note that for all  $u > 0$  and  $\omega \in \mathbb{R}$  we have  $\Phi_u(\omega) = \Phi_1(u^{1/p}\omega)$ . Hence, the function  $\Phi_u(\omega)$  can also be written as an expectation with respect to the random variable  $Z_1$  for all  $u > 0$  and  $\omega \in \mathbb{R}$ :

$$\Phi_u(\omega) = \mathbb{E}_{Z_1} \left[ e^{iu^{1/p}\omega Z_1} \right]. \quad (\text{B.1})$$

Equation (B.1) obviously also holds for  $u = 0$  since both sides of the equality are equal to 1 in that case. Using (B.1), we obtain

$$\mathbb{E}_{\mathbf{Z}_1} \left[ e^{iu^{1/p}\boldsymbol{\omega}^T \mathbf{Z}_1} \right] = e^{-u\|\boldsymbol{\omega}\|_p^p} \quad \forall u \geq 0, \quad (\text{B.2})$$

where  $\mathbf{Z}_1 = (Z_1^1, \dots, Z_1^D)^T$  is the concatenation of  $D$  independent copies of  $Z_1$ . We will use the latter representation in the second part of the proof.

*Proof that  $\varphi(\|\boldsymbol{\omega}\|_p^p)$  is a positive definite function.* We want to show that for any  $k \in \mathbb{N}$ , any  $\{\theta_\ell\}_{\ell=1}^k \subset \mathbb{R}^D$  and any  $\mathbf{c} \in \mathbb{C}^k \setminus \{\mathbf{0}_k\}$ , we have

$$\mathbf{c}^H \mathbf{G} \mathbf{c} > 0,$$

where  $(\cdot)^H$  denotes the conjugate transpose operator and

$$\mathbf{G}[\ell, \ell'] \triangleq \varphi(\|\theta_{\ell'} - \theta_\ell\|_p^p) \quad \forall \ell, \ell' \in \llbracket 1, k \rrbracket.$$

Note that in practice this will only be used for real-valued coefficients, but the result is established for complex-values  $\mathbf{c}$  to fit with the standard definition of a positive definite function. Since  $\varphi$  is a CMF, Lemma 2 ensures the existence of a non-negative finite Borel measure  $\nu$  such that for all  $\boldsymbol{\omega} \in \mathbb{R}^D$ :

$$\varphi(\|\boldsymbol{\omega}\|_p^p) = \int_0^{+\infty} e^{-u\|\boldsymbol{\omega}\|_p^p} d\nu(u).$$

Hence,

$$\begin{aligned} \mathbf{c}^H \mathbf{G} \mathbf{c} &= \sum_{\ell=1}^k \sum_{\ell'=1}^k \mathbf{c}[\ell]^* \mathbf{c}[\ell'] \varphi(\|\theta_{\ell'} - \theta_\ell\|_p^p) \\ &= \sum_{\ell=1}^k \sum_{\ell'=1}^k \mathbf{c}[\ell]^* \mathbf{c}[\ell'] \int_0^{+\infty} e^{-u\|\theta_{\ell'} - \theta_\ell\|_p^p} d\nu(u) \end{aligned} \quad (\text{B.3})$$

$$= \sum_{\ell=1}^k \sum_{\ell'=1}^k \mathbf{c}[\ell]^* \mathbf{c}[\ell'] \int_0^{+\infty} \mathbb{E}_{\mathbf{Z}_1} \left[ e^{iu^{1/p}(\theta_{\ell'} - \theta_\ell)^T \mathbf{Z}_1} \right] d\nu(u), \quad (\text{B.4})$$



where the last equality follows (B.2). By linearity of the expectation, it follows:

$$\begin{aligned}
\mathbf{c}^H \mathbf{G} \mathbf{c} &= \int_0^{+\infty} \mathbb{E}_{\mathbf{Z}_1} \left[ \sum_{\ell=1}^k \sum_{\ell'=1}^k \mathbf{c}[\ell]^* \mathbf{c}[\ell'] e^{-iu^{1/p}(\theta_\ell - \theta_{\ell'})^T \mathbf{Z}_1} \right] d\nu(u) \\
&= \int_0^{+\infty} \mathbb{E}_{\mathbf{Z}_1} \left[ \left( \sum_{\ell=1}^k \mathbf{c}[\ell] e^{iu^{1/p} \theta_\ell^T \mathbf{Z}_1} \right)^* \left( \sum_{\ell'=1}^k \mathbf{c}[\ell'] e^{iu^{1/p} \theta_{\ell'}^T \mathbf{Z}_1} \right) \right] d\nu(u) \\
&= \int_0^{+\infty} \mathbb{E}_{\mathbf{Z}_1} \left[ \left| \sum_{\ell=1}^k \mathbf{c}[\ell] e^{iu^{1/p} \theta_\ell^T \mathbf{Z}_1} \right|^2 \right] d\nu(u) \geq 0. \tag{B.5}
\end{aligned}$$

Since this holds for any  $\mathbf{c} \in \mathbb{C}^k$  this shows that  $\varphi(\|\cdot\|_p^p)$  is positive semi-definite.

To establish that  $\varphi(\|\cdot\|_p^p)$  is a positive definite function we now show that the equality in (B.5) can only occur if  $\mathbf{c} = \mathbf{0}_k$ . To this end, denote  $\psi(\mathbf{z}) \triangleq |\sum_{\ell=1}^k \mathbf{c}[\ell] e^{i\theta_\ell^T \mathbf{z}}|^2$  and  $\Psi(u) \triangleq \mathbb{E}_{\mathbf{Z}_1} [\psi(u^{1/p} \mathbf{Z}_1)]$  for  $u \in \mathbb{R}_+$ , and assume that equality holds in (B.5), that is to say  $\int_0^{+\infty} \Psi(u) d\nu(u) = 0$ . We next show that this implies that  $\mathbf{c} = \mathbf{0}_k$ .

First, we note that  $\Psi$  is continuous since

$$\begin{aligned}
\Psi(u) &= \mathbb{E}_{\mathbf{Z}_1} [\psi(u^{1/p} \mathbf{Z}_1)] \\
&= \sum_{\ell=1}^k \sum_{\ell'=1}^k \mathbf{c}[\ell]^* \mathbf{c}[\ell'] \mathbb{E}_{\mathbf{Z}_1} [e^{iu^{1/p}(\theta_{\ell'} - \theta_\ell)^T \mathbf{Z}_1}] \\
&= \sum_{\ell=1}^k \sum_{\ell'=1}^k \mathbf{c}[\ell]^* \mathbf{c}[\ell'] e^{-iu \|\theta_\ell - \theta_{\ell'}\|_p^p}
\end{aligned}$$

where the last equality follows (B.2).

Second, since  $\varphi$  is a CMF satisfying  $\varphi(0) = 1$  and  $\lim_{x \rightarrow \infty} \varphi(x) = 0$ , by Lemma 5 the non-negative finite Borel measure  $\nu$  satisfies  $\nu(\mathbb{R}_+^*) = 1$ . Since  $\mathbb{R}_+^* = \cup_{n \geq 1} [n, n+1] \cup \cup_{n \geq 1} [1/(n+1), 1/n]$ , there must exist  $n \geq 1$  such that either  $\nu([n, n+1]) > 0$  or  $\nu([1/(n+1), 1/n]) > 0$ . Without loss of generality consider the case  $\nu([n, n+1]) > 0$  (the other one can be treated similarly). Since  $\Psi$  is continuous over the compact set  $[n, n+1]$ , it attains its infimum over  $[n, n+1]$  at some  $u_0 \in [n, n+1]$ . Now, because  $\Psi(u) \geq 0$  for every  $u \geq 0$  we have

$$0 = \int_0^{+\infty} \Psi(u) d\nu(u) \geq \int_n^{n+1} \Psi(u) d\nu(u) \geq \Psi(u_0) \nu([n, n+1])$$

and we obtain  $\Psi(u_0) = 0$  since  $\nu([n, n+1]) > 0$ . By construction  $u_0 > 0$ .

Finally, we have that:

- the distribution of  $\mathbf{Z}_1$  has a density with respect to the Lebesgue measure and its density is continuous, except possibly at points where one coordinate vanishes.

- $\psi$  is nonnegative and continuous (by construction as the squared modulus of a finite linear combination of exponentials).

Hence, using the definition of  $\Psi(u_0) = \mathbb{E}_{\mathbf{Z}_1}[\psi(u_0^{1/p} \mathbf{Z}_1)]$ , we deduce that there exist  $\mathbf{z}_0 \in \mathbb{R}^D$  (with non-vanishing coordinates) and  $r > 0$  such that  $\psi(\mathbf{z}) = 0 \forall \mathbf{z} \in \mathcal{B}(\mathbf{z}_0, r)$ , where  $\mathcal{B}(\mathbf{z}_0, r)$  is the open ball of radius  $r$  centered at  $\mathbf{z}_0$ . For any  $\mathbf{y} \neq \mathbf{0}_D$  and  $0 \leq t \leq r/\|\mathbf{y}\|_2$ , we have  $\mathbf{z}_0 + t\mathbf{y} \in \mathcal{B}(\mathbf{z}_0, r)$  and therefore

$$\sum_{\ell=1}^k \mathbf{c}[\ell] e^{i\theta_\ell^T(\mathbf{z}_0 + t\mathbf{y})} = 0 \quad \forall t \in [0, r/\|\mathbf{y}\|_2]. \quad (\text{B.6})$$

If  $\mathbf{y}$  is such that  $\theta_\ell^T \mathbf{y} \neq \theta_{\ell'}^T \mathbf{y}$  for all  $\ell \neq \ell'$ , then the functions  $\{t \mapsto e^{it\theta_\ell^T \mathbf{y}}\}_{\ell=1}^k$  are linearly independent on  $[0, r/\|\mathbf{y}\|_2]$  and (B.6) holds if and only if  $\mathbf{c} = \mathbf{0}_k$ .

It thus remains to show that there exists some  $\mathbf{y} \in \mathbb{R}^D$  such that  $\theta_\ell^T \mathbf{y} \neq \theta_{\ell'}^T \mathbf{y}$  for all  $\ell \neq \ell'$ . To this end, let us consider the following finite set of vectors:

$$\mathcal{N} \triangleq \{\theta_\ell - \theta_{\ell'} : \ell, \ell' \in \llbracket 1, k \rrbracket, \ell \neq \ell'\}. \quad (\text{B.7})$$

As the parameters  $\theta_\ell$ 's are pairwise distinct, each  $\mathbf{n} \in \mathcal{N}$  is nonzero. Denote  $H_{\mathbf{n}}$  the linear hyperplane whose normal vector is  $\mathbf{n}$ , and consider  $H \triangleq \cup_{\mathbf{n} \in \mathcal{N}} H_{\mathbf{n}}$ . Since  $H$  is the union of a finite number of  $D$ -dimensional hyperplanes,  $\mathbb{R}^D \setminus H$  is not empty. Consider  $\mathbf{y} \in \mathbb{R}^D \setminus H$ . Then, by construction,  $\mathbf{n}^T \mathbf{y} \neq 0$  for all  $\mathbf{n} \in \mathcal{N}$  and therefore  $\theta_\ell^T \mathbf{y} \neq \theta_{\ell'}^T \mathbf{y}$  for all  $\ell \neq \ell'$ . This concludes the proof.

### B.2. Proof of Theorem 3

Our proof leverages Theorem 2 by showing that if  $\mathcal{A}$  is a CMF dictionary in dimension 1 with induced kernel  $\kappa$ , then:

- $\kappa$  is admissible in the sense of Definition 1,
- any finite support  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  is admissible with respect to kernel  $\kappa$  in the sense of Definition 2.

The result stated in Theorem 3 is then a direct consequence of Theorem 2. We begin with a more general lemma establishing the claim a):

**Lemma 6.** *Any CMF kernel  $\kappa \in \mathcal{K}_{\text{CMF}}(D)$ ,  $D \geq 1$ , is admissible.*

*Proof.* First, since  $\kappa \in \mathcal{K}_{\text{CMF}}(D)$ , there exists a CMF  $\varphi$  and scalar  $p \in ]0, 1]$  such that  $\varphi(0) = 1$  and  $\kappa(\theta, \theta') = \varphi(\|\theta - \theta'\|_p^p)$  for all  $\theta, \theta' \in \mathbb{R}^D$ . Hence  $\kappa(\theta, \theta) = \varphi(0) = 1$  for all  $\theta$ . Moreover, the function  $\theta' \mapsto \kappa(\theta, \theta')$  is continuous since both CMFs and  $\ell_p$ -norms are continuous. Hence  $\kappa$  satisfies (3.4). The fact that  $\kappa$  satisfies the vanishing property (3.6) is a straightforward consequence of the fact that  $\kappa(\theta, \theta') = \varphi(\|\theta - \theta'\|_p^p)$  and that  $\lim_{t \rightarrow +\infty} \varphi(t) = 0$ . Finally, we prove item ii) of Definition 1. As  $\varphi$  satisfies the assumptions of Lemma 5, it is strictly positive and strictly decreasing. This implies  $\kappa(\theta, \theta') > 0$  for any  $\theta, \theta'$ . Moreover, if  $\theta \neq \theta'$  then  $t = \|\theta - \theta'\|_p^p > 0$  and  $\varphi(t) < \varphi(0) = 1$ .  $\square$

The rest of this section is dedicated to the proof of claim *b*). Let us consider a non-empty subset of indices  $T \subseteq \llbracket 1, k \rrbracket$  with  $t = \text{card}(T)$ . Without loss of generality (up to some reordering), we assume that  $T = \llbracket 1, t \rrbracket$ . Let  $\{c_\ell\}_{\ell=1}^t \subset \mathbb{R}_+^*$  be such that

$$s \triangleq \sum_{\ell=1}^t c_\ell < 1 \quad (\text{B.8})$$

and consider the function

$$\begin{aligned} \psi: \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \sum_{\ell=1}^t c_\ell \kappa(\theta, \theta_\ell^*). \end{aligned} \quad (\text{B.9})$$

Using the integral formulation of CMF (see Lemma 2), we have that  $\psi$  is twice differentiable at any  $\theta \in \mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^t$  [72, proof of Th. 12a] and its second derivative writes:

$$\begin{aligned} \psi^{(2)}(\theta) &= \sum_{\ell=1}^t p(1-p)c_\ell \int_0^{+\infty} u |\theta_\ell^* - \theta|^{p-2} e^{-u|\theta_\ell^* - \theta|^p} d\nu(u) \\ &\quad + p^2 c_\ell \int_0^{+\infty} u^2 |\theta_\ell^* - \theta|^{2(p-1)} e^{-u|\theta_\ell^* - \theta|^p} d\nu(u) \end{aligned} \quad (\text{B.10})$$

We next show that items *i*) and *ii*) of Definition 2 hold.

*Item i) of Definition 2.* First, the vanishing property (3.6) of admissible kernels ensures that  $\psi$  admits at least one maximizer (see Lemma 1). We then show that any maximizer of  $\psi$  must necessarily belong to  $\{\theta_\ell^*\}_{\ell=1}^t$ .

Since  $\psi$  is twice continuously differentiable on  $\mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^t$ , any maximizer  $\theta_m \in \mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^t$  must verify the second-order optimality condition “ $\psi^{(2)}(\theta_m) \leq 0$ ”. Now this condition can never be fulfilled for  $\theta_m \in \mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^t$ . Indeed, we see that each integral term in (B.10) is positive since  $\theta_m \notin \{\theta_\ell^*\}_{\ell=1}^t$  and, by Lemma 5,  $\nu(\mathbb{R}_+^*) > 0$ . Since  $p \in ]0, 1]$  and  $c_\ell > 0 \forall \ell$ , it follows that  $\psi^{(2)}(\theta_m) > 0$ .

*Item ii) of Definition 2.* Assume  $T = \llbracket 1, t \rrbracket \neq \llbracket 1, k \rrbracket$  and consider  $t' \in \llbracket 1, k \rrbracket \setminus T$ . Without loss of generality, suppose that  $t' = t + 1$  and consider

$$\begin{aligned} \phi: \mathbb{R} &\longrightarrow \mathbb{R} \\ \theta &\longmapsto \psi(\theta) - \kappa(\theta, \theta_{t+1}^*). \end{aligned} \quad (\text{B.11})$$

From the definition of  $\psi$  in (B.9),  $\phi$  writes

$$\phi(\theta) = \sum_{\ell=1}^t c_\ell \varphi\left(|\theta - \theta_\ell^*|_p^p\right) - \varphi\left(|\theta - \theta_{t+1}^*|_p^p\right). \quad (\text{B.12})$$

Assume that  $\phi(\theta_\ell^*) \leq 0$  for every  $\ell \in \llbracket 1, t \rrbracket$ . We will then show that  $\phi(\theta) \leq 0$  for every  $\theta \in \mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^t$  so that item *ii*) of Definition 2 is satisfied.

Suppose there exists  $\theta_0 \in \mathbb{R}$  such that  $\phi(\theta_0) > 0$  and let us show that this leads to a contradiction.

Let us first emphasize that the existence of some  $\theta_0 \in \mathbb{R}$  such that  $\phi(\theta_0) > 0$  implies that a maximizer of  $\phi$  exists. Indeed, we have that  $\lim_{\theta \rightarrow \pm\infty} \phi(\theta) = 0$  since

$$|\phi(\theta)| \leq \sum_{\ell=1}^t c_\ell \kappa(\theta, \theta_\ell^*) + \kappa(\theta, \theta_{t+1}^*) \quad (\text{B.13})$$

and  $\kappa$  obeys the vanishing property (3.6) by hypothesis. Hence, for any  $0 < \varepsilon < \phi(\theta_0)$  there exists a compact set  $K_\varepsilon$  such that

$$\forall \theta \in K_\varepsilon^c, \quad \phi(\theta) \leq \varepsilon < \sup_{\theta' \in K_\varepsilon} \phi(\theta'), \quad (\text{B.14})$$

because  $\phi$  is continuous. The extreme value theorem [73, Prop. A.8] then states that (at least) one maximizer of  $\phi$ , say  $\theta_m$ , exists and  $\theta_m \in K_\varepsilon$ . We note that by definition  $\phi(\theta_m) \geq \phi(\theta_0) > 0$ . We show below that we also must necessarily have  $\phi(\theta_m) \leq 0$ . This leads to the desired contradiction and proves the result.

Let  $\lambda_\ell \triangleq |\theta_\ell^* - \theta_m|$  and assume without loss of generality that

$$\lambda_1 \leq \dots \leq \lambda_t. \quad (\text{B.15})$$

We have that  $\lambda_1 > 0$  because  $\theta_m \notin \{\theta_\ell^*\}_{\ell=1}^t$  since  $\phi(\theta_\ell^*) \leq 0 \forall \ell \in \llbracket 1, t \rrbracket$  (by assumption) and  $\phi(\theta_m) \geq \phi(\theta_0) > 0$ . We next show that the working assumptions also imply  $\phi(\theta_m) \leq 0$  by distinguishing between three cases:

**Case 1:**  $\lambda_{t+1} \leq \lambda_1$ . From (B.15), we have:

$$\forall u \geq 0, \quad \max_{1 \leq \ell \leq t} e^{-u\lambda_\ell^p} \leq e^{-u\lambda_{t+1}^p}. \quad (\text{B.16})$$

Hence

$$\phi(\theta_m) \leq \underbrace{\left( \sum_{\ell=1}^t c_\ell - 1 \right)}_{<0 \text{ by hyp.}} \underbrace{\int_0^{+\infty} e^{-u\lambda_{t+1}^p} d\nu(u)}_{>0} < 0.$$

**Case 2:**  $\lambda_{t+1} > \lambda_t$ . We rely on the following technical lemma that exploits the notion of “sign changes of a finite sequence”. This notion is defined as the number of times two consecutive elements of the finite sequence have opposite signs. For instance, the sequence  $(1, 1, -1, 1)$  has two sign changes (respectively at the third and fourth positions).

**Lemma 7.** *Let  $P(u) \triangleq \sum_{\ell=1}^k c_\ell e^{-\lambda_\ell u}$  be an exponential polynomial on  $\mathbb{R}_+$  with  $0 < \lambda_1 < \dots < \lambda_k$  and  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$ . Assume that:*

- *the sequence  $c_1, \dots, c_k$  has at most two sign changes;*
- *$P(0) < 0$  and  $\lim_{u \rightarrow +\infty} P(u) = 0_+$ .*

Then there exists  $u_0 > 0$  for which the following inequality holds

$$\int_0^{+\infty} f(u)P(u) d\nu(u) \geq f(u_0) \int_0^{+\infty} P(u) d\nu(u), \quad (\text{B.17})$$

for any non-decreasing function  $f$  on  $\mathbb{R}_+$  and any (unsigned) finite Borel measure  $\nu$  on  $\mathbb{R}_+$  such that the integrals converge.

The proof of the lemma is postponed to Appendix C.2.

As mentioned previously, we have on the one hand that  $\theta_m \notin \{\theta_\ell^*\}_{\ell=1}^t$ . On the other hand,  $\theta_m \neq \theta_{t+1}^*$  because  $|\theta_m - \theta_{t+1}^*| = \lambda_{t+1} > \lambda_1 > 0$ . Therefore,  $\theta_m \in \mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^{t+1}$ . Since  $\phi$  is twice continuously differentiable on  $\mathbb{R} \setminus \{\theta_\ell^*\}_{\ell=1}^{t+1}$ ,  $\theta_m$  must necessarily verify the following second-order optimality condition:

$$\phi^{(2)}(\theta_m) \leq 0. \quad (\text{B.18})$$

We next show that  $C\phi(\theta_m) < \phi^{(2)}(\theta_m)$  for some positive constant  $C > 0$ . Hence, in view of (B.18), this leads to the desired contradiction:  $\phi(\theta_m) < 0$ .

Assume first that  $\lambda_1 < \dots < \lambda_t$  (the equality cases will be addressed later). From (B.10) we have that  $\phi^{(2)} = \phi_1^{(2)} + \phi_2^{(2)}$  with

$$\begin{aligned} \phi_1^{(2)}(\theta_m) &= p(1-p) \int_0^{+\infty} u \left( \sum_{\ell=1}^t c_\ell \lambda_\ell^{p-2} e^{-u\lambda_\ell^p} - \lambda_{t+1}^{p-2} e^{-u\lambda_{t+1}^p} \right) d\nu(u) \\ \phi_2^{(2)}(\theta_m) &= p^2 \int_0^{+\infty} u^2 \left( \sum_{\ell=1}^t c_\ell \lambda_\ell^{2(p-1)} e^{-u\lambda_\ell^p} - \lambda_{t+1}^{2(p-1)} e^{-u\lambda_{t+1}^p} \right) d\nu(u). \end{aligned}$$

Using  $\lambda_{t+1} > \lambda_\ell > 0 \forall \ell \in \llbracket 1, t \rrbracket$ , we obtain

$$\begin{aligned} \phi_2^{(2)}(\theta_m) &= \frac{p^2}{\lambda_{t+1}^{2(1-p)}} \int_0^{+\infty} u^2 \left( \sum_{\ell=1}^t c_\ell \left( \frac{\lambda_{t+1}}{\lambda_\ell} \right)^{2(1-p)} e^{-u\lambda_\ell^p} - e^{-u\lambda_{t+1}^p} \right) d\nu(u) \\ &> \frac{p^2}{\lambda_{t+1}^{2(1-p)}} \int_0^{+\infty} u^2 \left( \sum_{\ell=1}^t c_\ell e^{-u\lambda_\ell^p} - e^{-u\lambda_{t+1}^p} \right) d\nu(u). \end{aligned} \quad (\text{B.19})$$

Note now that:

- the function  $u \mapsto u^2$  is increasing,
- $P(u) \triangleq \sum_{\ell=1}^t c_\ell e^{-u\lambda_\ell^p} - e^{-u\lambda_{t+1}^p}$  is an exponential polynomial with  $0 < \lambda_1 < \dots < \lambda_t < \lambda_{t+1}$  and whose sequence of coefficients is  $(c_1, \dots, c_t, -1)$  and has exactly one sign change.
- As  $\max_{1 \leq \ell \leq t} \lambda_\ell < \lambda_{t+1}$  by hypothesis, we have  $P(u) > 0$  for sufficiently large  $u$  so  $\lim_{u \rightarrow +\infty} P(u) = 0^+$ .
- Since  $\sum_{\ell=1}^t c_\ell < 1$  we have  $P(0) < 0$ .

Therefore, Lemma 7 applies and there exists  $u_0 > 0$  such that

$$\phi_2^{(2)}(\theta_m) > \frac{p^2}{\lambda_{t+1}^{2(1-p)}} u_0^2 \int_0^{+\infty} P(u) d\nu(u) = \frac{p^2}{\lambda_{t+1}^{2(1-p)}} u_0^2 \phi(\theta_m). \quad (\text{B.20})$$

This establishes that  $\phi_2^{(2)}(\theta_m) > C_2 \phi(\theta_m)$  where  $C_2 > 0$  is a positive constant. The same rationale leads to  $\phi_1^{(2)}(\theta_m) > C_1 \phi(\theta_m)$  with  $C_1 \geq 0$  ( $C_1 = 0$  for  $p = 1$  since  $\phi_1^{(2)}$  is identically zero). Since  $\phi^{(2)} = \phi_1^{(2)} + \phi_2^{(2)}$ , one obtains that  $\phi^{(2)}(\theta_m) > (C_1 + C_2) \phi(\theta_m)$ , which concludes the proof for  $\lambda_1 < \dots < \lambda_t$ .

Let us now come back to the general case where  $\lambda_1 \leq \dots \leq \lambda_t$ . Denote by  $\tilde{\lambda}_1 < \dots < \tilde{\lambda}_{t'}$ , with  $t' \leq t$ , the ordered distinct values in  $\{\lambda_\ell\}_{\ell=1}^t$ , and let  $\tilde{\lambda}_{t'+1} = \lambda_{t+1}$ . Moreover, for any  $\ell' \in \llbracket 1, t' \rrbracket$ , let  $\tilde{c}_{\ell'}$  be equal to the sum of the coefficients  $c_\ell$  such that  $\lambda_\ell = \tilde{\lambda}_{\ell'}$ , and let  $\tilde{c}_{t'+1} = -1$ . We note that, by definition,  $\sum_{\ell'=1}^{t'} \tilde{c}_{\ell'} = \sum_{\ell=1}^t c_\ell < 1$ . We can then show that  $\phi^{(2)}(\theta_m) > C \phi(\theta_m)$  with  $C > 0$  by applying the same reasoning as above to  $\tilde{\lambda}_1, \dots, \tilde{\lambda}_{t'+1}$  and  $\tilde{c}_1, \dots, \tilde{c}_{t'+1}$ .

**Case 3:**  $\lambda_1 < \lambda_{t+1} \leq \lambda_t$ . There exists  $\ell_0 \in \llbracket 1, t-1 \rrbracket$  such that:

$$\begin{cases} \lambda_\ell < \lambda_{t+1} & \text{for } \ell \leq \ell_0 \\ \lambda_\ell \geq \lambda_{t+1} & \text{for } \ell > \ell_0. \end{cases} \quad (\text{B.21a})$$

$$\quad \quad \quad (\text{B.21b})$$

Denote  $\varepsilon \triangleq 1 - \sum_{\ell=1}^t c_\ell > 0$  and let  $s_1 \triangleq \sum_{\ell=1}^{\ell_0} c_\ell + \frac{\varepsilon}{2}$  and  $s_2 \triangleq \sum_{\ell=\ell_0+1}^t c_\ell + \frac{\varepsilon}{2}$  such that  $s_1 + s_2 = 1$ . One can write

$$\begin{aligned} \phi(\theta_m) &= s_1 \underbrace{\int_0^{+\infty} \left( \sum_{\ell=1}^{\ell_0} \frac{c_\ell}{s_1} e^{-u\lambda_\ell^p} - e^{-u\lambda_{t+1}^p} \right) d\nu(u)}_{\triangleq \phi_1(\theta_m)} \\ &\quad + s_2 \underbrace{\int_0^{+\infty} \left( \sum_{\ell=\ell_0+1}^t \frac{c_\ell}{s_2} e^{-u\lambda_\ell^p} - e^{-u\lambda_{t+1}^p} \right) d\nu(u)}_{\triangleq \phi_2(\theta_m)}. \end{aligned} \quad (\text{B.22})$$

Using (B.21a) and the fact that  $\sum_{\ell=1}^{\ell_0} \frac{c_\ell}{s_1} < 1$ , we have that  $\phi_1(\theta_m) < 0$  by resorting to the same reasoning as in Case 2. Similarly, using (B.21b) and the fact that  $\sum_{\ell=\ell_0+1}^t \frac{c_\ell}{s_2} < 1$ , we obtain that  $\phi_2(\theta_m) < 0$  by the same arguments as in Case 1. Hence, we finally have  $\phi(\theta_m) \leq s_1 \phi_1(\theta_m) + s_2 \phi_2(\theta_m) < 0$ , which leads to the desired contradiction.

### B.3. Proof of Theorem 4 - Recovery in dimension $D$

Let  $\mathcal{S}^* = \{\theta_\ell^*\}_{\ell=1}^k$  and assume  $\text{Grid}(\mathcal{S}^*)$  is axis admissible. Let us first observe that, by virtue of Corollary 1, the elements of  $\{\mathbf{a}(\theta) : \theta \in \text{Grid}(\mathcal{S}^*)\}$  are linearly independent. Let

$$\mathcal{R} \triangleq \text{span}(\{\mathbf{a}(\theta) : \theta \in \text{Grid}(\mathcal{S}^*)\}). \quad (\text{B.23})$$

If  $\mathbf{r} \in \mathcal{R} \setminus \{\mathbf{0}_{\mathcal{H}}\}$ , we note from Lemma 1 that the function  $f : \theta \mapsto |\langle \mathbf{a}(\theta), \mathbf{r} \rangle|$  admits at least one maximizer since  $\mathbf{r}$  results from a (non trivial) finite linear combination of atoms. Moreover, the maximum of  $f$  must be strictly greater than zero. If not,  $\langle \mathbf{a}(\theta), \mathbf{r} \rangle = 0 \ \forall \theta \in \text{Grid}(\mathcal{S}^*)$  and one deduces that  $\mathbf{r} \in \mathcal{R}^\perp$ , the orthogonal to  $\mathcal{R}$ . Since  $\mathbf{r} \in \mathcal{R}$ , this leads to  $\mathbf{r} = \mathbf{0}_{\mathcal{H}}$  which is in contradiction with our initial assumption “ $\mathbf{r} \neq \mathbf{0}_{\mathcal{H}}$ ”. We also have that the maximizers of  $f$  must belong to  $\text{Grid}(\mathcal{S}^*)$  as shown by the following arguments.

If  $\theta_m$  is a maximizer of  $f$ , then  $t = \theta_m[d]$  is a maximizer of

$$f_d : t \mapsto |\langle \mathbf{a}(\theta_m + (t - \theta_m[d])\mathbf{e}_d), \mathbf{r} \rangle|. \quad (\text{B.24})$$

Denoting  $\theta_0 \triangleq \theta_m - \theta_m[d]\mathbf{e}_d$ , we have  $\theta_0 \perp \mathbf{e}_d$  by construction and  $f_d$  writes

$$f_d(t) = \left| \sum_{\ell=1}^k c_\ell \kappa(\theta_0 + t\mathbf{e}_d, \theta_\ell^*) \right|. \quad (\text{B.25})$$

Because  $\text{Grid}(\mathcal{S}^*)$  is axis admissible with respect to  $\kappa$  that  $f_d$  is not identically zero (since  $f_d(\theta_m[d]) = f(\theta_m) > 0$ ), the maximizers of  $f_d$  must belong to  $\mathcal{S}_d \triangleq \{\theta_\ell^*[d]\}_{\ell=1}^k$ . As a consequence, since this conclusion holds for any  $d \in \llbracket 1, D \rrbracket$ , we have that  $\theta_m \in \prod_{d=1}^D \mathcal{S}_d = \text{Grid}(\mathcal{S}^*)$ . Formulated in a slightly different way, we thus just proved that:

$$\forall \mathbf{r} \in \mathcal{R} \setminus \{\mathbf{0}_{\mathcal{H}}\}, \forall \theta \in \Theta \setminus \text{Grid}(\mathcal{S}^*), \max_{\theta' \in \text{Grid}(\mathcal{S}^*)} |\langle \mathbf{a}(\theta'), \mathbf{r} \rangle| > |\langle \mathbf{a}(\theta), \mathbf{r} \rangle|. \quad (\text{B.26})$$

We are now ready to prove the statements of the theorem:

*Grid( $\mathcal{S}^*$ )-delayed recovery of any  $\mathcal{S} \subseteq \mathcal{S}^*$ .* First, since  $\mathcal{S} \subseteq \text{Grid}(\mathcal{S}^*)$ , we have that  $\mathbf{y} \in \mathcal{R}$ . Moreover,  $\mathbf{y} \neq \mathbf{0}_{\mathcal{H}}$  since it results from a nontrivial linear combination of linearly independent atoms. Hence (B.26) holds and OMP selects a parameter in  $\text{Grid}(\mathcal{S}^*)$  at the first iteration. Repeating the same argument at the next iterations, OMP selects parameters in  $\text{Grid}(\mathcal{S}^*)$  until the residual  $\mathbf{r}$  vanishes. Now, because the atoms in  $\{\mathbf{a}(\theta) : \theta \in \text{Grid}(\mathcal{S}^*)\}$  are linearly independent,  $\mathbf{r} = \mathbf{0}_{\mathcal{H}}$  if and only if the set of parameters selected by OMP, say  $\widehat{\mathcal{S}}$ , verifies  $\mathcal{S} \subseteq \widehat{\mathcal{S}}$ . Since OMP never selects twice the same parameter and  $\widehat{\mathcal{S}} \subseteq \text{Grid}(\mathcal{S}^*)$ , we thus achieve  $\text{Grid}(\mathcal{S}^*)$ -delayed recovery of  $\mathcal{S}$ .

*$\mathcal{S}^*$ -delayed recovery of any  $\mathcal{S} \subseteq \mathcal{S}^*$ .* Since the elements of  $\{\mathbf{a}(\theta) : \theta \in \text{Grid}(\mathcal{S}^*)\}$  are linearly independent, (3.30-R-ERC) can equivalently be rewritten as (see e.g., [1, Prop. 3.15]):

$$\forall \mathbf{r} \in \mathcal{R}_{\mathcal{S}^*} \setminus \{\mathbf{0}_{\mathcal{H}}\}, \forall \theta \in \text{Grid}(\mathcal{S}^*) \setminus \mathcal{S}^*, \max_{\theta' \in \mathcal{S}^*} |\langle \mathbf{a}(\theta'), \mathbf{r} \rangle| > |\langle \mathbf{a}(\theta), \mathbf{r} \rangle|,$$

where  $\mathcal{R}_{\mathcal{S}^*} \triangleq \text{span}(\{\mathbf{a}(\theta_\ell^*)\}_{\ell=1}^k)$ . Combining this result with (B.26) and using the fact that  $\mathcal{R}_{\mathcal{S}^*} \subset \mathcal{R}$  lead to

$$\forall \mathbf{r} \in \mathcal{R}_{\mathcal{S}^*} \setminus \{\mathbf{0}_{\mathcal{H}}\}, \forall \theta \in \Theta \setminus \mathcal{S}^*, \max_{\theta' \in \mathcal{S}^*} |\langle \mathbf{a}(\theta'), \mathbf{r} \rangle| > |\langle \mathbf{a}(\theta), \mathbf{r} \rangle|.$$

Following the same arguments as above, we then have that: 1) OMP selects parameters in  $\mathcal{S}^*$  until the residual  $\mathbf{r}$  vanishes; 2)  $\mathbf{r} = \mathbf{0}_{\mathcal{H}}$  if and only if the set of parameters selected by OMP verifies  $\mathcal{S} \subseteq \hat{\mathcal{S}}$ . Since OMP never selects twice the same parameter and  $\hat{\mathcal{S}} \subseteq \mathcal{S}^*$ , OMP thus achieves  $\mathcal{S}^*$ -delayed recovery of  $\mathcal{S}$ .

*Sharpness of the result.* If (3.30-R-ERC) is not verified, we have from [1, Prop. 3.15] that there exists some  $\mathbf{y}_{\text{bad}} \in \mathcal{R}_{\mathcal{S}^*} \setminus \{\mathbf{0}_{\mathcal{H}}\}$  such that

$$\max_{\theta \in \text{Grid}(\mathcal{S}^*) \setminus \mathcal{S}^*} |\langle \mathbf{a}(\theta), \mathbf{y}_{\text{bad}} \rangle| > \max_{\theta \in \mathcal{S}^*} |\langle \mathbf{a}(\theta), \mathbf{y}_{\text{bad}} \rangle|. \quad (\text{B.27})$$

In other words, OMP with  $\mathbf{y}_{\text{bad}}$  as input selects some  $\theta \notin \mathcal{S}^*$  at the first iteration.

#### B.4. Proof of Lemma 4

Let  $\mathcal{G} = \prod_{d=1}^D \mathcal{S}_d = \{\theta_\ell\}_{\ell=1}^{\text{card}(\mathcal{G})}$  be an arbitrary Cartesian grid in  $\mathbb{R}^D$  and  $\{c_\ell\}_{\ell=1}^{\text{card}(\mathcal{G})} \subset \mathbb{R}$  be a set of  $\text{card}(\mathcal{G})$  coefficients not all equal to 0. Consider  $d \in \llbracket 1, D \rrbracket$  and  $\theta_0 \in \mathbb{R}^D$  such that  $\theta_0[d] = 0$  and define

$$f_d: \mathbb{R} \longrightarrow \mathbb{R}_+ \\ t \longmapsto \left| \sum_{\ell=1}^{\text{card}(\mathcal{G})} c_\ell \kappa(\theta_0 + t\mathbf{e}_d, \theta_\ell) \right| = \left| \sum_{\ell=1}^{\text{card}(\mathcal{G})} c_\ell e^{-\lambda \|\theta_0 + t\mathbf{e}_d - \theta_\ell\|_p^p} \right|. \quad (\text{B.28})$$

We assume that  $f_d$  is not identically zero as in the statement of Definition 7.  $f_d$  can be rewritten as

$$f_d(t) = \left| \sum_{\ell=1}^{\text{card}(\mathcal{G})} c_\ell e^{-\lambda |t - \theta_\ell[d]|^p - \lambda \sum_{j=1, j \neq d}^D |\theta_0[j] - \theta_\ell[j]|^p} \right|.$$

Let  $q$  denote the number of distinct elements of  $\{\theta_\ell[d]\}_{\ell=1}^{\text{card}(\mathcal{G})}$  and suppose (up to some renumbering) that  $\{\theta_\ell[d]\}_{\ell=1}^q$  are pairwise distinct. We note that  $q \geq 1$  because otherwise there is a contradiction with our hypothesis “ $f_d$  not identically zero”. We can then rewrite  $f_d$  as

$$f_d(t) = \left| \sum_{\ell=1}^q \tilde{c}_\ell e^{-\lambda |t - \theta_\ell[d]|^p} \right| \quad (\text{B.29})$$

where the terms proportional to  $e^{|t - \theta_\ell[d]|^p}$  for (possibly) identical values of  $\theta_\ell[d]$  have been merged together, and the scalars  $\tilde{c}_\ell$  take into account the constant terms in the exponentials that do not depend on  $t$ .

Let  $\mathcal{A}_1 = \{\mathbf{a}_1(t') : t' \in \mathbb{R}\}$  be a Generalized Laplace dictionary in dimension 1 (see Definition 5). Then,  $f_d(t)$  can also be interpreted as the inner product between atom  $\mathbf{a}_1(t) \in \mathcal{A}_1$  and  $\mathbf{y}_1 \triangleq \sum_{\ell=1}^q \tilde{c}_\ell \mathbf{a}_1(\theta_\ell[d])$ . Let  $\tilde{\mathcal{S}} \triangleq \{\theta_\ell[d] : \tilde{c}_\ell \neq 0, \ell = 1 \dots q\}$ . Applying Theorem 3, we have that OMP with  $\mathbf{y}_1$  as input



achieves exact  $\text{card}(\tilde{\mathcal{S}})$ -step recovery of  $\tilde{\mathcal{S}}$ . In particular, this implies that OMP selects a parameter in  $\tilde{\mathcal{S}}$  at the first iteration, that is:

$$\forall t' \in \mathbb{R} \setminus \tilde{\mathcal{S}}, \quad f_d(t') < \max_{t \in \tilde{\mathcal{S}}} f_d(t). \quad (\text{B.30})$$

Hence, the maximizers of  $f_d$  belong to  $\tilde{\mathcal{S}} \subseteq \{\theta_\ell[d]\}_{\ell=1}^q \subseteq \mathcal{S}_d$ .

## C. Miscellaneous

### C.1. Proof of Lemma 1

Let  $\mathbf{r} \in \mathcal{H}$ ,  $k \in \mathbb{N}$  and assume that there exists  $k$  parameters  $\{\theta_\ell^*\}_{\ell=1}^k \subset \Theta$  and  $k$  nonzero coefficients  $\{c_\ell\}_{\ell=1}^k \subset \mathbb{R}^*$  such that  $\mathbf{r} = \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$ . Define function  $\varphi$  as

$$\begin{aligned} \varphi: \Theta &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto |\langle \mathbf{a}(\theta), \mathbf{r} \rangle| = \left| \sum_{\ell=1}^k c_\ell \kappa(\theta, \theta_\ell^*) \right|, \end{aligned} \quad (\text{C.1})$$

that is, the function involved in step 6 in Algorithm 1. We now prove that a maximizer of (C.1) exists. To that aim, we distinguish two cases.

*Case 1:*  $\forall \theta \in \Theta, \varphi(\theta) = 0$ . In that case, any parameter  $\theta \in \Theta$  is a maximizer of  $\varphi$ .

*Case 2:*  $\exists \theta_0 \in \Theta, \varphi(\theta_0) > 0$ . Denote  $\varepsilon \triangleq \varphi(\theta_0)$ . We then have

$$\sup_{\theta \in \Theta} \varphi(\theta) \geq \varphi(\theta_0) = \varepsilon > 0. \quad (\text{C.2})$$

Hence, by condition (3.6), there exists  $k$  compact sets  $\{K_\ell\}_{\ell=1}^k$  such that for all  $\ell \in \llbracket 1, k \rrbracket$ ,  $\theta_0 \in K_\ell$  and

$$\forall \theta \in K_\ell^c, \quad \kappa(\theta, \theta_\ell^*) < \frac{1}{\sum_{\ell'=1}^k |c_{\ell'}|} \varepsilon. \quad (\text{C.3})$$

Note that the right-hand-side of (C.3) is well defined: by positive-definiteness of  $\langle \cdot, \cdot \rangle$ ,  $\mathbf{r} \neq 0$  so  $k > 0$  and  $\sum_{\ell=1}^k |c_\ell| > 0$  necessarily. Define  $K = \cup_{\ell=1}^k K_\ell$ . Since  $K^c = \cap_{\ell=1}^k K_\ell^c$ , we have using the triangular inequality

$$\forall \theta \in K^c, \quad \varphi(\theta) \leq \sum_{\ell=1}^k |c_\ell| \kappa(\theta, \theta_\ell^*) < \varepsilon. \quad (\text{C.4})$$

See now that  $\varphi$  is continuous by continuity of  $\kappa$ ,  $K$  is compact as a union of compact sets. Then, the extreme value theorem ensures that there exists  $\theta_m$  such that  $\varphi(\theta_m) \geq \varphi(\theta)$  for all  $\theta \in K$ . Lemma 1 follows by seeing that  $\varphi(\theta_m) \geq \varphi(\theta)$  for all  $\theta$  by (C.4).

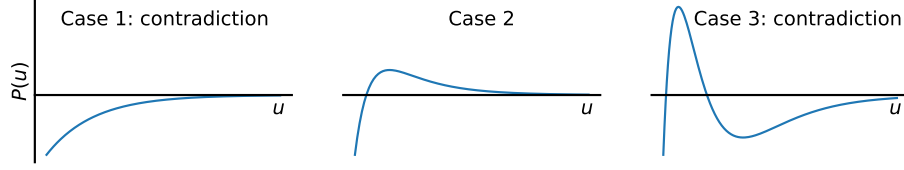


Figure C.2: Shape of  $P$  (see proof of Lemma 7) with constraints *i*)  $P$  is continuous, *ii*)  $P(u) < 0$  and *iii*)  $\exists u_0 > 0$  such that  $P(u) > 0$  for all  $u > u_0$ . One see that the constraints cannot be satisfied in cases 1 and 3.

### C.2. Proof of Lemma 7

The proof of Lemma 7 is based on the following result:

**Lemma 8** (Laguerre's generalization of Descartes's rule of signs [74, p. 319]). *Let  $a_1, \dots, a_k$  be nonzero real coefficients and  $0 < x_1 < \dots < x_k$  be real numbers. Let  $z$  be the number of real roots of the function  $P(u) = \sum_{\ell=1}^k a_\ell x_\ell^u$ , and  $n_c$  be the number of changes in sign in the sequence of numbers  $a_1, \dots, a_k$ . Then  $z \leq n_c$ .*

The sequence of coefficients  $a_\ell = c_{k+1-\ell}$  with  $\ell \in \llbracket 1, k \rrbracket$  has only two sign changes by hypothesis. By applying Lemma 8 with  $x_\ell = e^{-\lambda_{k+1-\ell}}$ , one sees that  $P$  has at most two real roots, so at most two sign changes on  $\mathbb{R}_+$ . However,  $P$  must satisfy the following constraints:

- i*)  $P$  is continuous on  $[0, +\infty[$ ,
- ii*)  $P(0) < 0$ ,
- iii*) there exists  $u_0 > 0$  such that  $P(u) > 0$  for all  $u > u_0$ .

As illustrated in Figure C.2, these three constraints cannot be verified simultaneously if  $P$  has 0 or 2 roots. Thus  $P$  has exactly one sign change on  $\mathbb{R}_+$  and there exists  $u_0 > 0$  such that  $u < u_0 \implies P(u) < 0$  and  $u > u_0 \implies P(u) > 0$ . One then has, for any non-decreasing function  $f$  and any (non-negative) measure  $\nu$  on  $\mathbb{R}_+$ :

$$\begin{aligned}
 & \int f(u)P(u) \, d\nu(u) \\
 &= \int_0^{u_0} \underbrace{f(u)}_{\text{non-decreasing}} \underbrace{P(u)}_{\leq 0} \, d\nu(u) + \int_{u_0}^{+\infty} \underbrace{f(u)}_{\text{non-decreasing}} \underbrace{P(u)}_{\geq 0} \, d\nu(u) \\
 &\geq \int_0^{u_0} f(u_0)P(u) \, d\nu(u) + \int_{u_0}^{+\infty} f(u_0)P(u) \, d\nu(u) \\
 &= f(u_0) \int_0^{+\infty} P(u) \, d\nu(u).
 \end{aligned}$$

#### D. Details related to Example 4

Assume that  $\varphi$  is right differentiable at 0. We first prove by contradiction that  $\varphi^{(1)}(0) < 0$ . Assume that  $\varphi^{(1)}(0) = 0$ . As  $\varphi$  is a CMF,  $\varphi^{(1)}(t) \leq 0$  for all  $t \in \mathbb{R}_+$  and  $\varphi^{(1)}$  is non-decreasing on  $\mathbb{R}_+$ . It follows that  $\varphi^{(1)}$  is identically zero. Hence  $\varphi$  is constant and equal to  $\varphi(0) = 1$  which contradicts the assumption  $\lim_{t \rightarrow +\infty} \varphi(t) = 0$ .

Consider now the function defined for all  $x \geq 0$  by  $f : x \mapsto (k-1)\varphi(2x) - k\varphi(x) + 1$ . We note that  $f$  corresponds to the quantity involved in (3.24) with the substitution  $x \leftrightarrow \Delta^p$ . Since  $\varphi(0) = 1$ , we have  $f(0) = 0$ . Moreover  $f$  is differentiable for any  $x > 0$  and

$$\begin{aligned} f^{(1)}(x) &= 2(k-1)\varphi^{(1)}(2x) - k\varphi^{(1)}(x) \\ &= k\varphi^{(1)}(x) \left[ 2\left(1 - \frac{1}{k}\right) \frac{\varphi^{(1)}(2x)}{\varphi^{(1)}(x)} - 1 \right]. \end{aligned} \quad (\text{D.1})$$

Since  $\varphi$  is right differentiable at 0, the ratio  $\frac{\varphi^{(1)}(2x)}{\varphi^{(1)}(x)}$  tends to 1 as  $x$  tends to 0 (remember that we proved that  $\varphi^{(1)}(0) < 0$ ). Since  $k \geq 3$  we have  $2(1 - \frac{1}{k}) > 1$ , hence there exists  $x_0 > 0$  such that

$$x < x_0 \Rightarrow 2\left(1 - \frac{1}{k}\right) \frac{\varphi^{(1)}(2x)}{\varphi^{(1)}(x)} - 1 > 0. \quad (\text{D.2})$$

By Lemma 5, we have that  $\varphi^{(1)}(x) < 0$  for all  $x > 0$ . Hence  $f^{(1)}(x) < 0$  for  $x < x_0$ , that is  $f$  is decreasing on  $[0, x_0]$ . Combining this result with  $f(0) = 0$ , we deduce that (3.24) holds whenever  $\Delta^p < x_0$ , that is the wrong parameter  $\mathbf{0}_D$  will be preferred to any of the  $\{\theta_\ell^*\}_{\ell=1}^k$ .

#### E. Exact recovery in higher dimensions - CMF kernel and $k = 2$

In this section, we elaborate on the notion of “axis admissibility” (see Definition 7) for general CMF kernels. We first show (see Example 5) that there exist some Cartesian grids which are not axis admissible with respect to some CMF kernels. We then emphasize in the case  $k = 2$  that the notion of “axis admissibility” is not necessary to achieve  $k$ -step recovery in CMF dictionaries.

**Example 5.** Let  $\mathcal{A}$  be a CMF dictionary with induced kernel  $\kappa = \varphi(\|\cdot - \cdot\|_p^p)$  and consider  $\Delta > 0$ ,  $c_1 = c_2 = 1$ ,  $c_3 = c_4 = -\frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)}$  and

$$\forall t \in \mathbb{R}, \quad f_1(t) = \left| \sum_{\ell=1}^4 c_\ell \kappa(t\mathbf{e}_1, \theta_\ell) \right| \quad (\text{E.1})$$

where  $\mathbf{e}_d$  the  $d$ -th canonical basis vector of  $\mathbb{R}^2$ .

Let  $\mathcal{G} \triangleq \{\theta_\ell\}_{\ell=1}^4 \subset \mathbb{R}^2$  be a Cartesian grid with  $\theta_1 = (0, 0) = \mathbf{0}_2$ ,  $\theta_2 = (\Delta, 0) = \Delta\mathbf{e}_1$ ,  $\theta_3 = (0, \Delta) = \Delta\mathbf{e}_2$ ,  $\theta_4 = (\Delta, \Delta) = \Delta\mathbf{1}_2$ . Simple algebraic manipulations then show that  $f_1(0) = f_1(\Delta) = 0$  and

$$f_1\left(\frac{\Delta}{2}\right) = 2\varphi\left(\frac{1}{2^p}\Delta^p\right) \left| 1 - \frac{1+\varphi(\Delta^p)}{\varphi(\Delta^p)+\varphi(2\Delta^p)} \frac{\varphi(\Delta^p + \frac{1}{2^p}\Delta^p)}{\varphi(\frac{1}{2^p}\Delta^p)} \right|. \quad (\text{E.2})$$

If  $\varphi$  and  $\Delta$  are such that  $f_1(\frac{\Delta}{2}) \neq 0$ , one can conclude that the maximizers of  $f_1$  are distinct from 0 and  $\Delta$ . In view of Definition 7, this shows that  $\mathcal{G}$  is not axis admissible with respect to  $\mathcal{A}$ .

For instance, this is the case for the CMF  $\varphi : x \mapsto \frac{1}{1+x}$  (cf Example 2). Indeed, in this case (E.2) particularizes to

$$f_1(\frac{\Delta}{2}) = \frac{2}{1 + \frac{\Delta^p}{2^p}} \left| 1 - \frac{2 + \Delta^p}{1 + \frac{1+\Delta^p}{1+2\Delta^p}} \frac{1 + \frac{\Delta^p}{2^p}}{1 + \Delta^p + \frac{\Delta^p}{2^p}} \right|. \quad (\text{E.3})$$

As the factor inside the absolute value in the right-hand side is a non-zero rational function of  $x = \Delta^p$ , we have  $f_1(\Delta/2) \neq 0$  except possibly on a set of values of  $\Delta$  which has Lebesgue measure equal to zero. Hence there exists  $\Delta > 0$  such that  $f_1(\Delta/2) > 0$ .

We finally note that the construction presented here for the case  $D = 2$  easily extends to  $D > 2$  by zero-padding of the  $\theta_\ell$ 's.

We next show that  $k$ -step recovery of  $\mathcal{S}^*$  with  $k = \text{card}(\mathcal{S}^*) = 2$  may be possible in CMF dictionaries even when the axis admissibility assumption fails to hold. First, we state and prove a useful technical lemma:

**Lemma 9.** *Let  $\kappa$  be a CMF kernel in dimension  $D$  in the sense of Definition 4. For any  $\theta_1, \theta_2, \theta_3 \in \mathbb{R}^D$ , the following result holds:*

$$\kappa(\theta_1, \theta_2) \kappa(\theta_2, \theta_3) \leq \kappa(\theta_1, \theta_3). \quad (\text{E.4})$$

*Proof.* By definition, there exists a CMF  $\varphi$  such that  $\kappa(\cdot, \cdot) = \varphi(\|\cdot - \cdot\|_p^p)$  and  $\varphi(0) = 1$ . Since  $\varphi$  is nonnegative and decreasing, we have for all  $x, y \geq 0$

$$\varphi(x+y) \geq \varphi(x+y)\varphi(x+y) \geq \varphi(x)\varphi(y). \quad (\text{E.5})$$

Using this result with  $x = \|\theta_1 - \theta_2\|_p^p$  and  $y = \|\theta_2 - \theta_3\|_p^p$ , we have

$$\kappa(\theta_1, \theta_2) \kappa(\theta_2, \theta_3) \leq \varphi(\|\theta_1 - \theta_2\|_p^p + \|\theta_2 - \theta_3\|_p^p). \quad (\text{E.6})$$

Since the quasi-norm  $\|\cdot\|_p^p$  satisfies a triangular inequality, we have  $\|\theta_1 - \theta_3\|_p^p \leq \|\theta_1 - \theta_2\|_p^p + \|\theta_2 - \theta_3\|_p^p$ . As any CMF is decreasing, (E.4) follows.  $\square$

We are now ready to state our recovery result:

**Lemma 10** (Exact recovery for CMF dictionaries when  $k = 2$ ). *Let  $\mathcal{A}$  be a CMF dictionary in dimension  $D \geq 1$  with induced kernel  $\kappa$ . Consider a support  $\mathcal{S}^* = \{\theta_1^*, \theta_2^*\}$  where  $\theta_1^* \neq \theta_2^*$ , and let  $\mathbf{G} \in \mathbb{R}^{2 \times 2}$  be the matrix defined by  $\mathbf{G}[\ell, \ell'] = \kappa(\theta_\ell^*, \theta_{\ell'}^*)$ . Assume that*

$$\forall \theta \in \text{Grid}(\mathcal{S}^*) \setminus \mathcal{S}^*, \quad \|\mathbf{G}^{-1} \mathbf{g}_\theta\|_1 < 1 \quad (\text{E.7})$$

where  $\mathbf{g}_\theta \in \mathbb{R}^2$  is defined by  $\mathbf{g}_\theta[\ell] = \kappa(\theta, \theta_\ell^*)$  for  $\ell = 1, 2$ . Then OMP achieves exact 2-step recovery of  $\mathcal{S}^*$ .

*Proof.* By Lemma 6,  $\kappa$  is admissible in the sense of Definition 1. We show below that since (E.7) holds,  $\mathcal{S}^*$  is admissible with respect to  $\kappa$  in the sense of Definition 2. Lemma 10 then follows from Theorem 2.

Consider a non-empty subset of indices  $T \subseteq \{1, 2\}$  and  $t \triangleq \text{card}(T)$ . Let also  $\{c_\ell\}_{\ell \in T}$  be such that  $c_\ell > 0$  and  $\sum_{\ell \in T} c_\ell < 1$ . Define

$$\begin{aligned} \psi: \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ \theta &\longmapsto \sum_{\ell \in T} c_\ell \kappa(\theta, \theta_\ell^*) \end{aligned} \quad (\text{E.8})$$

We next show that items *i)* and *ii)* of Definition 2 are satisfied.

*Item i) of Definition 2.* We distinguish two cases:

- If  $t = 1$ , we can assume without loss of generality that  $T = \{1\}$ . Since  $\kappa(\theta, \theta_1^*) < 1$  for all  $\theta \neq \theta_1^*$ , one immediately sees that  $\psi(\theta) = c_1 \kappa(\theta, \theta_1^*) < c_1 = \psi(\theta_1^*)$  for all  $\theta \neq \theta_1^*$ . Hence,  $\theta_1^*$  is the unique global maximizer of  $\psi$ .
- If  $t = 2$ , let  $\theta_m$  be a maximizer of  $\psi$ . We note that  $\psi$  can also be written as  $\psi(\theta) = |\langle \mathbf{a}(\theta), \mathbf{y} \rangle|$  where  $\mathbf{y} \triangleq \sum_{\ell=1}^k c_\ell \mathbf{a}(\theta_\ell^*)$  hence a maximizer always exists by virtue of Lemma 1. Since  $\theta_m$  maximizes the  $D$ -dimensional function  $\psi$ , its  $d$ -th entry  $\theta_m[d]$  is a maximizer of the one-dimensional section of  $\psi$  along the  $d$ -th canonical direction, denoted  $\psi_d$ :

$$\begin{aligned} \psi_d: \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto \sum_{\ell \in T} c_\ell \varphi(|x - \theta_\ell^*[d]|^p + \sum_{j \neq d} |\theta_m[j] - \theta_\ell^*[j]|^p). \end{aligned} \quad (\text{E.9})$$

Applying the same reasoning as in the proof of Theorem 3 (see part of the proof dedicated to establishing “item *i)* of Definition 2”), we have  $\forall x \notin \{\theta_\ell^*[d]\}_{\ell=1}^2$ :  $\psi_d$  is twice differentiable and  $\psi_d^{(2)}(x) > 0$ . Hence, no  $x \in \{\theta_\ell^*[d]\}_{\ell=1}^2$  can be a maximizer and necessarily  $\theta_m[d] \in \{\theta_\ell^*[d]\}_{\ell=1}^2$ . Since this result is valid for all  $d \in \llbracket 1, D \rrbracket$ , we finally have  $\theta_m \in \text{Grid}(\mathcal{S}^*)$ .

Therefore, since (E.7) holds, we have

$$\begin{aligned} \max_{\theta \in \text{Grid}(\mathcal{S}^*) \setminus \mathcal{S}^*} \psi(\theta) &= \max_{\theta \in \text{Grid}(\mathcal{S}^*) \setminus \mathcal{S}^*} |\langle \mathbf{a}(\theta), \mathbf{y} \rangle| \\ &< \max_{\theta^* \in \mathcal{S}^*} |\langle \mathbf{a}(\theta_\ell^*), \mathbf{y} \rangle| = \max_{\theta^* \in \mathcal{S}^*} \psi(\theta_\ell^*) \end{aligned} \quad (\text{E.10})$$

Hence all maximizers of  $\psi$  belong to  $\mathcal{S}^*$ .

*Item ii) of Definition 2.* From the working assumptions of item *ii)*, the set  $T$  satisfies  $T \neq \emptyset$  and there exists  $\ell \in \{1, 2\} \setminus T$ . Hence, we have  $T \neq \{1, 2\}$ , that is  $T$  is a singleton. We assume without loss of generality that  $T = \{1\}$ . Hence  $\psi(\theta) = c_1 \kappa(\theta, \theta_1^*)$  for some  $0 < c_1 < 1$ . If  $\psi_1(\theta_1^*) - \kappa(\theta_1^*, \theta_2^*) \leq 0$ , then  $c_1 - \kappa(\theta_1^*, \theta_2^*) = \psi_1(\theta_1^*) - \kappa(\theta_1^*, \theta_2^*) \leq 0$ . Hence  $c_1 \leq \kappa(\theta_1^*, \theta_2^*)$  and  $\psi(\theta) \leq \kappa(\theta_1^*, \theta_2^*) \kappa(\theta, \theta_1^*)$  for each  $\theta$ . Using Lemma 9 with  $\theta_1 = \theta$ ,  $\theta_2 = \theta_1^*$ ,  $\theta_3 = \theta_2^*$ , we obtain for each  $\theta \in \Theta$ :

$$\psi(\theta) - \kappa(\theta, \theta_2^*) \leq \kappa(\theta, \theta_1^*) \kappa(\theta_1^*, \theta_2^*) - \kappa(\theta, \theta_2^*) \leq 0.$$

□

## F. Table of notations

NOTATION	COMMENT
<i>General notations</i>	
$\mathcal{H}, \mathbf{y}$	(Hilbert) observation space and observation
$\mathcal{A}, \mathbf{a}(\cdot)$	Dictionary $\mathcal{A}$ made of parametric atoms $\mathbf{a}$
$\mu$	Coherence between atoms of a support
$\mathbf{c} \in \mathbb{R}^k$	Weighting coefficients
$\Theta, \theta$	Parameter set and element
$\mathcal{S}, \mathcal{S}^*$	Set of parameters
$\mathcal{G}$	Cartesian grid
$k, \ell$	Number of atoms, most frequent index
Grid	Set augmenter, see (3.27)
$\varphi$	CMF (see Definition 3)
$\kappa$	Kernel function $\Theta \times \Theta \rightarrow \mathbb{R}_+$
$\mathcal{K}_{\text{CMF}}(D)$	Set of CMF kernels in dimension $D$
$\mathcal{K}_{\text{Lap}}(D)$	Set of Laplace kernels in dimension $D$
$f^{(n)}$	$n$ -th derivative of function $f$
$\mathbf{e}_\ell$	$\ell$ -th element of the canonical basis
$\mathbb{E}$	Expectation operator
$\mathbf{i}$	Imaginary number
<i>Technical notations</i>	
$\mathbf{G}, \mathbf{g}_\ell$	Gram matrix related to a support $\mathcal{S}$ , columns of $\mathbf{G}$
$\mathbf{g}_\theta$	parametric vector related to a support $\mathcal{S}$
$\mathbf{u}, \mathbf{v}$	Vector of $\mathbb{R}^k$ for some $k$ often defined as $\mathbf{u}, \mathbf{v} = \mathbf{G}^{-1} \mathbf{g}_\theta$ for some $\theta \in \Theta$

Table F.1: Table of notations.

## Acknowledgments

Part of this work has been funded thanks to the Becose ANR project no. ANR-15-CE23-0021.

- [1] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser Basel, 2013. doi:[10.1007/978-0-8176-4948-7](https://doi.org/10.1007/978-0-8176-4948-7).
- [2] C. Ekanadham, D. Tranchina, E. P. Simoncelli, Recovery of sparse translation-invariant signals with continuous basis pursuit, *IEEE Transactions on Signal Processing* 59 (10) (2011) 4735–4744. doi:[10.1109/TSP.2011.2160058](https://doi.org/10.1109/TSP.2011.2160058).
- [3] E. J. Candès, C. Fernandez-Granda, Towards a Mathematical Theory of Super-resolution, *Communications on Pure and Applied Mathematics* 67 (6) (2014) 906–956. doi:[10.1002/cpa.21455](https://doi.org/10.1002/cpa.21455).
- [4] V. Duval, G. Peyré, Exact support recovery for sparse spikes deconvolution, *Foundations of Computational Mathematics* 15 (5) (2014) 1315–1355. doi:[10.1007/s10208-014-9228-6](https://doi.org/10.1007/s10208-014-9228-6).
- [5] Y. C. Pati, R. Rezaeiifar, P. S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1. doi:[10.1109/ACSSC.1993.342465](https://doi.org/10.1109/ACSSC.1993.342465).
- [6] K. S. Miller, S. G. Samko, Completely monotonic functions, *Integral Transforms and Special Functions* 12 (4) (2001) 389–402.
- [7] B. Natarajan, Sparse approximate solutions to linear systems, *SIAM Journal on Computing* 24 (2) (1995) 227–234. doi:[10.1137/S0097539792240406](https://doi.org/10.1137/S0097539792240406).
- [8] A. J. Miller, *Subset selection in regression*, Chapman and Hall, London, 2002.
- [9] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing* 41 (12) (1993) 3397–3415. doi:[10.1109/78.258082](https://doi.org/10.1109/78.258082).
- [10] R. A. DeVore, V. N. Temlyakov, Some remarks on greedy algorithms, *Advances in Computational Mathematics* 5 (1) (1996) 173–187. doi:[10.1007/bf02124742](https://doi.org/10.1007/bf02124742).
- [11] E. Liu, V. N. Temlyakov, The Orthogonal Super Greedy Algorithm and Applications in Compressed Sensing, *IEEE Transactions on Information Theory* 58 (4) (2012) 2040–2047. doi:[10.1109/TIT.2011.2177632](https://doi.org/10.1109/TIT.2011.2177632).
- [12] S. Chen, S. A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, *International Journal of Control* 50 (5) (1989) 1873–1896. doi:[10.1080/00207178908953472](https://doi.org/10.1080/00207178908953472).

- [13] J. H. Friedman, W. Stuetzle, Projection pursuit regression, *Journal of the American Statistical Association* 76 (376) (1981) 817–823. doi:[10.1080/01621459.1981.10477729](https://doi.org/10.1080/01621459.1981.10477729).
- [14] P. J. Huber, Projection pursuit, *The Annals of Statistics* 13 (2) (1985) 435–475. doi:[10.1214/aos/1176349519](https://doi.org/10.1214/aos/1176349519).
- [15] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, *IEEE Signal Processing Letters* 9 (4) (2002) 137–140. doi:[10.1109/LSP.2002.1001652](https://doi.org/10.1109/LSP.2002.1001652).
- [16] V. N. Temlyakov, Greedy approximation, *Acta Numerica* 17 (2008) 235–409. doi:[10.1017/s0962492906380014](https://doi.org/10.1017/s0962492906380014).
- [17] P. Vincent, Y. Bengio, Kernel Matching Pursuit, *Machine Learning* 48 (1/3) (2002) 165–187. doi:[10.1023/a:1013955821559](https://doi.org/10.1023/a:1013955821559).
- [18] J. F. Claerbout, F. Muir, Robust modeling with erratic data, *Geophysics* 38 (5) (1973) 826–844. doi:[10.1190/1.1440378](https://doi.org/10.1190/1.1440378).
- [19] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* 20 (1) (1998) 33–61. doi:[10.1137/S1064827596304010](https://doi.org/10.1137/S1064827596304010).
- [20] R. Tibshirani, [Regression shrinkage and selection via the lasso](#), *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1) (1996) 267–288.  
URL <http://www.jstor.org/stable/2346178>
- [21] R. Tibshirani, I. Johnstone, T. Hastie, B. Efron, Least angle regression, *The Annals of Statistics* 32 (2) (2004) 407–499. doi:[10.1214/009053604000000067](https://doi.org/10.1214/009053604000000067).
- [22] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences* 2 (1) (2009) 183–202. doi:[10.1137/080716542](https://doi.org/10.1137/080716542).
- [23] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine Learning* 3 (1) (2011) 1–122. doi:[10.1561/22000000016](https://doi.org/10.1561/22000000016).
- [24] R. Gribonval, M. Nielsen, Beyond sparsity: Recovering structured representations by  $\ell^1$  minimization and greedy algorithms, *Advances in Computational Mathematics* 28 (1) (2008) 23–41. doi:[10.1007/s10444-005-9009-5](https://doi.org/10.1007/s10444-005-9009-5).
- [25] L. Borup, R. Gribonval, M. Nielsen, Beyond coherence: Recovering structured time-frequency representations, *Applied and Computational Harmonic Analysis* 24 (1) (2008) 120–128. doi:[10.1016/j.acha.2007.09.002](https://doi.org/10.1016/j.acha.2007.09.002).



- [26] K. C. Knudson, J. Yates, A. Huk, J. W. Pillow, Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit, *Advances in Neural Information Processing Systems* 27 (2014) 1215–1223.
- [27] A. Eftekhari, M. B. Wakin, Greed is super: A fast algorithm for super-resolution (2015). [arXiv:1511.03385](https://arxiv.org/abs/1511.03385).
- [28] C. Dorffer, A. Drémeau, C. Herzet, [Efficient atom selection strategy for iterative sparse approximations](#), in: *iTWIST 2018 - International Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques*, Marseille, France, 2018, pp. 1–3. URL <https://hal.inria.fr/hal-01937501>
- [29] K. Bredies, H. K. Pikkariainen, Inverse problems in spaces of measures, *ESAIM: Control, Optimisation and Calculus of Variations* 19 (1) (2012) 190–218. [doi:10.1051/cocv/2011205](https://doi.org/10.1051/cocv/2011205).
- [30] Y. de Castro, F. Gamboa, Exact reconstruction using Beurling minimal extrapolation, *Journal of Mathematical Analysis and Applications* 395 (1) (2012) 336 – 354. [doi:10.1016/j.jmaa.2012.05.011](https://doi.org/10.1016/j.jmaa.2012.05.011).
- [31] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, *IEEE Transactions on Information Theory* 59 (11) (2013) 7465–7490. [doi:10.1109/tit.2013.2277451](https://doi.org/10.1109/tit.2013.2277451).
- [32] Y. D. Castro, F. Gamboa, D. Henrion, J.-B. Lasserre, Exact solutions to super resolution on semi-algebraic domains in higher dimensions, *IEEE Transactions on Information Theory* 63 (1) (2017) 621–630. [doi:10.1109/tit.2016.2619368](https://doi.org/10.1109/tit.2016.2619368).
- [33] N. Boyd, G. Schiebinger, B. Recht, The alternating descent conditional gradient method for sparse inverse problems, in: *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015, pp. 57–60. [doi:10.1109/CAMSAP.2015.7383735](https://doi.org/10.1109/CAMSAP.2015.7383735).
- [34] P. Catala, V. Duval, G. Peyré, A low-rank approach to off-the-grid sparse deconvolution, *SIAM Journal on Imaging Sciences* 12 (3) (2019) 1464–1500. [doi:10.1137/19M124071X](https://doi.org/10.1137/19M124071X).
- [35] Q. Denoyelle, V. Duval, G. Peyré, E. Soubies, The sliding frank–wolfe algorithm and its application to super-resolution microscopy, *Inverse Problems* 36 (1) (2019) 014001. [doi:10.1088/1361-6420/ab2a29](https://doi.org/10.1088/1361-6420/ab2a29).
- [36] A. Eftekhari, A. Thompson, Sparse inverse problems over measures: Equivalence of the conditional gradient and exchange methods, *SIAM Journal on Optimization* 29 (2) (2019) 1329–1349. [doi:10.1137/18m1183388](https://doi.org/10.1137/18m1183388).
- [37] A. Flinth, F. de Gournay, P. Weiss, On the linear convergence rates of exchange and continuous methods for total variation minimization (Jun 2019). [arXiv:1906.09919](https://arxiv.org/abs/1906.09919).

- [38] L. Chizat, F. Bach, On the global convergence of gradient descent for over-parameterized models using optimal transport, in: *Advances in Neural Information Processing Systems* 31, 2018, pp. 3036–3046.
- [39] L. Chizat, *Sparse Optimization on Measures with Over-parameterized Gradient Descent*, working paper or preprint (July 2019).  
URL <https://hal.archives-ouvertes.fr/hal-02190822>
- [40] G. de Prony, Essai experimental et analytique : sur les lois de la dilatabilite des fluides elastique et sur celles de la force expansive de la vapeur de l’eau et de la vapeur de l’alkool, a differentes temperatures, 1795, journal de l’Ecole Polytechnique.
- [41] S. Kunis, T. Peter, T. Römer, U. von der Ohe, A multivariate generalization of Prony’s method, *Linear Algebra and its Applications* 490 (2016) 31–47. doi:10.1016/j.laa.2015.10.023.
- [42] W. Liao, A. Fannjiang, MUSIC for single-snapshot spectral estimation: Stability and super-resolution, *Applied and Computational Harmonic Analysis* 40 (1) (2016) 33–67. doi:10.1016/j.acha.2014.12.003.
- [43] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37 (7) (1989) 984–995. doi:10.1109/29.32276.
- [44] X. Wei, P. L. Dragotti, FRESH—FRI-based single-image super-resolution algorithm, *IEEE Transactions on Image Processing* 25 (8) (2016) 3723–3735. doi:10.1109/tip.2016.2563178.
- [45] J. A. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Transactions on Information Theory* 50 (10) (2004) 2231–2242. doi:10.1109/TIT.2004.834793.
- [46] C. Soussen, R. Gribonval, J. Idier, C. Herzet, Joint  $k$ -Step Analysis of Orthogonal Matching Pursuit and Orthogonal Least Squares, *IEEE Transactions on Information Theory* 59 (5) (2013) 3158–3174. doi:10.1109/tit.2013.2238606.
- [47] R. Gribonval, P. Vandergheynst, On the exponential convergence of matching pursuits in quasi-incoherent dictionaries, *IEEE Transactions on Information Theory* 52 (1) (2006) 255–261. doi:10.1109/tit.2005.860474.
- [48] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Transactions on Information Theory* 50 (6) (2004) 1341–1344. doi:10.1109/TIT.2004.828141.
- [49] J. A. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, *IEEE Transactions on Information Theory* 52 (3) (2006) 1030–1051. doi:10.1109/TIT.2005.864420.

- [50] C. Herzet, A. Drémeau, C. Soussen, Relaxed Recovery Conditions for OMP/OLS by Exploiting Both Coherence and Decay, *IEEE Transactions on Information Theory* 62 (1) (2016) 459–470. doi:[10.1109/TIT.2015.2490660](https://doi.org/10.1109/TIT.2015.2490660).
- [51] S. Huang, J. Zhu, Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP, *Inverse Problems* 27 (3) (2011) 035003. doi:[10.1088/0266-5611/27/3/035003](https://doi.org/10.1088/0266-5611/27/3/035003).
- [52] R. Maleh, Improved RIP Analysis of Orthogonal Matching Pursuit, Tech. rep. (2011). [arXiv:1102.4311](https://arxiv.org/abs/1102.4311).
- [53] Q. Mo, Y. Shen, A Remark on the Restricted Isometry Property in Orthogonal Matching Pursuit, *IEEE Transactions on Information Theory* 58 (6) (2012) 3654–3656. doi:[10.1109/TIT.2012.2185923](https://doi.org/10.1109/TIT.2012.2185923).
- [54] J. Wang, B. Shim, On the Recovery Limit of Sparse Signals Using Orthogonal Matching Pursuit, *IEEE Transactions on Signal Processing* 60 (9) (2012) 4973–4976. doi:[10.1109/TSP.2012.2203124](https://doi.org/10.1109/TSP.2012.2203124).
- [55] L. Chang, J. Wu, An Improved RIP-Based Performance Guarantee for Sparse Signal Recovery via Orthogonal Matching Pursuit, *IEEE Transactions on Information Theory* 60 (9) (2014) 5702–5715. doi:[10.1109/TIT.2014.2338314](https://doi.org/10.1109/TIT.2014.2338314).
- [56] J. Wen, X. Zhu, D. Li, Improved bounds on restricted isometry constant for orthogonal matching pursuit, *Electronics Letters* 49 (23) (2013) 1487–1489. doi:[10.1049/el.2013.2222](https://doi.org/10.1049/el.2013.2222).
- [57] Q. Mo, A Sharp Restricted Isometry Constant Bound of Orthogonal Matching Pursuit, Tech. rep. (2015). [arXiv:1501.01708](https://arxiv.org/abs/1501.01708).
- [58] W. Rudin, *Principles of Mathematical Analysis*, 3rd Edition, International Series in Pure and Applied Mathematics, McGraw-Hill Science/Engineering/Math, 1976.
- [59] D. D. Carlo, C. Elvira, A. Deleforge, N. Bertin, R. Gribonval, Blaster: An off-grid method for blind and regularized acoustic echoes retrieval, in: *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 156–160.
- [60] J.-M. Azaïs, Y. de Castro, F. Gamboa, Spike detection from inaccurate samplings, *Applied and Computational Harmonic Analysis* 38 (2) (2015) 177–195. doi:[10.1016/j.acha.2014.03.004](https://doi.org/10.1016/j.acha.2014.03.004).
- [61] C. Fernandez-Granda, Super-resolution of point sources via convex programming, *Information and Inference: A Journal of the IMA* 5 (3) (2016) 251–303. doi:[10.1093/imaiai/iaw005](https://doi.org/10.1093/imaiai/iaw005).

- [62] M. Kreĭn, A. Nudel'man, The Markov Moment Problem and Extremal Problems, American Mathematical Society, 1977. doi:[10.1090/mmono/050](https://doi.org/10.1090/mmono/050).
- [63] Q. Denoyelle, V. Duval, G. Peyré, Support Recovery for Sparse Super-Resolution of Positive Measures, Journal of Fourier Analysis and Applications 23 (5) (2017) 1153–1194.
- [64] C. Poon, G. Peyré, MultiDimensional sparse super-resolution, SIAM Journal on Mathematical Analysis 51 (1) (2019) 1–44. doi:[10.1137/17m1147822](https://doi.org/10.1137/17m1147822).
- [65] C. J. Hillar, L.-H. Lim, Most tensor problems are np-hard, J. ACM 60 (6). doi:[10.1145/2512329](https://doi.org/10.1145/2512329).  
URL <https://doi.org/10.1145/2512329>
- [66] C. Elvira, J. E. Cohen, C. Herzet, R. Gribonval, Continuous dictionaries meet low-rank tensor approximations, in: iTwist 2020 - International Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques, Nantes, France, 2020, pp. 1–3.  
URL <https://hal.archives-ouvertes.fr/hal-02567115>
- [67] V. Chandrasekaran, B. Recht, P. A. Parrilo, A. S. Willsky, The convex geometry of linear inverse problems, Foundations of Computational Mathematics 12 (6) (2012) 805–849. doi:[10.1007/s10208-012-9135-7](https://doi.org/10.1007/s10208-012-9135-7).  
URL <https://doi.org/10.1007/s10208-012-9135-7>
- [68] H. Wendland, Scattered data approximation, 2005. doi:[10.2277/0521843359](https://doi.org/10.2277/0521843359).
- [69] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, USA, 2004.
- [70] D. Bernstein, Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory, Princeton University Press, 2005.
- [71] G. Pólya, Remarks on Characteristic Functions, in: Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Calif., 1949, pp. 115–123.  
URL <https://projecteuclid.org/euclid.bsmsp/1166219202>
- [72] D. V. Widder, Laplace Transform, hardcover Edition, Princeton Mathematical Press, 1941.
- [73] D. Bertsekas, Nonlinear Programming, 2nd Edition, Athena Scientific, 1999.
- [74] H. Fejzić, C. Freiling, D. Rinne, Descartes' rule of signs, alternations of data sets, and balanced differences, The American Mathematical Monthly 116 (4) (2009) 316–327.  
URL <http://www.jstor.org/stable/40391091>